A Publication of the American Psychological Association

The Journal of Abnormal and Social Psychology

Founded by Morton Prince, 1906

CONTENTS

ARTICLES

Development and Validation of a Projective Measure of Power Motivation . . 1
Joseph Veroff

GSR Reactivity as a Function of Anxiety, Instructions, and Sex 9
James L. Berry and Barclay Martin

The Relationships Among Imaginative, Direct Verbal, and Physiological Measures of Anxiety in an Achievement Situation 13
Alfred C. Raphelson

Conditioning of a Response Class on a Personality Test 19
Anne M. Vaihmann

The Generality of Attitudes Toward Authority and Nonauthority Figures . . . 24
Leroy S. Barwen and Donald T. Campbell

Guessing Behavior and Autism 32
Charles M. Solley, Douglas N. Jackson, and Samuel J. Messick

Accuracy of Interpersonal Prediction as a Function of Judge and Object Characteristics 37
Bela O. Baker and Jack Block

Differences Between Schizophrenic and Brain-Damaged Groups in Conceptual Aspects of Object Sorting 44
Laurence S. McGaughran and Louis J. Moruss

A Comparison Between Ipsative and Normative Ratings of Personality 50
Jack Block

Continued on inside front cover

Volume 54 JANUARY 1957 Number 1

CONTENTS—*Continued*

The Effects of Group Expectations and Self-Esteem Upon Self-Evaluation . . 55
Ezra Stotland, Stanley Thorley, Edwin Thomas, Arthur S. Cohen, and Alvin Zander

Age Differences in Verbal Learning 64
Sheldon J. Korchin and Harold Basowitz

Personality, Problem Solving, and the Einstellung Effect 79
Brendan A. Maher

On the Relation Between the Probability of a Word as an Association and in General Linguistic Usage 85
Davis Howes

Looking at Occupations 86
Willa Freeman Grimes

Age Differences in the Perception of Closure 93
Harold Basowitz and Sheldon J. Korchin

Postdecision Exposure to Relevant Information 98
Danuta Ehrlich, Isaiah Guttman, Peter Schönbach, and Judson Mills

A New Look at Vigilance and Defense 103
Donald P. Spence

Practice Effects in Reaction-Time Tasks in Brain-Injured Patients 109
Arthur L. Benton and Harold L. Blackburn

A Test of the Taylor-Spence Theory of Anxiety 114
Eli Salis and Arthur J. Hoehn

Breadth of Deviate Concepts Used by Schizophrenics 118
Loren J. Chapman and Janet A. Taylor

CRITIQUE AND NOTES

Effect of Preinformation Upon Human Characteristics Attributed to Animated Geometric Figures 124
Ronald E. Shor

The Effect of Stimulus Class on the Problem-Solving Behavior of Schizophrenics and Normals 126
Robert H. Davis and Robert W. Harrington

Response Set in the F Scale 129
Loren J. Chapman and Donald T. Campbell

A Note on "Ethnocentrism" and Acquiescent Response Sets 132
Douglas N. Jackson and Samuel J. Messick

A Note on "Discrimination Reaction Time as a Function of Anxiety and Intelligence" 134
James T. Freeman and Howard Maher

Social Status and Abstract Behavior 135
Donald C. Findlay and Carson McGuire

How "Rigid" is the "Authoritarian?" 137
Douglas N. Jackson, Samuel J. Messick, and Charles M. Solley

Anality, Aggression, and Acquiescence 140
William Robinowitz

Volume 54 JANUARY, 1957 Number 1

The Journal of Abnormal and Social Psychology

Editor: M. BREWSTER SMITH, New York University

Associate Editor: E. J. SHOBEN, JR., Teachers College, Columbia University

CONSULTING EDITORS

GORDON W. ALLPORT
Harvard University

URIE BRONFENBRENNER
Cornell University

DONALD T. CAMPBELL
Northwestern University

LAUNOR F. CARTER
The RAND Corporation

RICHARD CHRISTIE
Columbia University

MORTON DEUTSCH
Bell Telephone Laboratories

I. E. FARBER
State University of Iowa

LEON FESTINGER
Stanford University

DONALD W. FISKE
University of Chicago

ROBERT R. HOLT
New York University

HERBERT H. HYMAN
Columbia University

IRVING L. JANIS
Yale University

HAROLD E. JONES
University of California

HAROLD H. KELLEY
Center for Advanced Study in the Behavioral Sciences

LEONARD S. KOGAN
Community Service Society of New York

GARDNER LINDZEY
Syracuse University

DAVID C. MCCLELLAND
Harvard University

QUINN MCNEMAR
Stanford University

LEO POSTMAN
University of California

NEVITT SANFORD
Vassar College

SEYMOUR SARASON
Yale University

RICHARD L. SOLOMON
Harvard University

RENATO TAGIURI
Harvard University

CENTRAL OFFICE EDITORIAL STAFF

Managing Editor: ARTHUR C. HOFFMAN

Assistant Managing Editor: HELEN ORR

Editorial Staff: FRANCES HAZEL; SADIE J. DOYLE

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

MT. ROYAL & GUILFORD AVES., BALTIMORE 2, MD.

AND 1333 SIXTEENTH STREET N.W., WASHINGTON 6, D. C.

Entered as second-class matter February 15, 1956 at the Post Office at Baltimore, Maryland under the Act of February 28, 1925. Acceptance for mailing at special rate of postage provided for in Section 34.40 (e), P. L. and R. Additional entry at Washington, D.C.

Copyright 1957 by the American Psychological Association, Inc.

The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis; those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area, this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior

The JOURNAL is issued bimonthly in January, March, May, July, September, and November, two volumes per year. The two volumes average 784 pages. The subcription price per year is $16.00; foreign, $16.50. Single issues are $3.00.

Original manuscripts should be sent to the Editor:

DR. M. BREWSTER SMITH
Graduate Department of Psychology
New York University
Washington Square
New York 3, N. Y.

Fifty free offprints are supplied to contributors. Authors of prior publication article receive no gratis offprints.

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street N.W.
Washington 6, D. C.

Changes of address must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Printed and mailed by:

WAVERLY PRESS, INC.
Baltimore 2, Maryland

DEVELOPMENT AND VALIDATION OF A PROJECTIVE MEASURE OF POWER MOTIVATION[1]

JOSEPH VEROFF

Survey Research Center, University of Michigan

THE purpose of this study was to develop a valid measure of power motivation. McClelland and his associates have shown that scoring stories written in response to pictures can be used as a basis for measures of strength of the achievement (8) and the affiliation (3, 11) motives. These investigators have shown that with an increase in experimentally aroused achievement or affiliation motivation, there was in the stories an increase in the frequency of certain kinds of imagery related to these motives. Furthermore, the coding categories, which can be used to reflect these shifts in imagery can be successfully employed as a basis for measuring individual differences in achievement and affiliation motivations. Subjects who have high scores on these measures are significantly different from subjects who have low scores in behaviors which are meaningfully related to the strength of the motives.

Following the same general procedures of these studies, the experimenter anticipated that a parallel measure of another motive—the power motive—could be developed.

What is meant by power motivation? Numerous social theorists have used a concept of power. In some cases, the concept is relegated to the part it plays in the analysis of social structure. In other cases, social power is a dimension to be considered in the analysis of interpersonal relationships. In the latter sense, the power motive will be considered that disposition directing behavior toward satisfactions contingent upon the control of the means of influencing another person(s). In the phenomenal sphere of the power-motivated individual, he considers himself the "gatekeeper" to certain decision-making of others. The means of control can be anything at all that can be used to manipulate another person. Overt dominance strivings can be considered one kind of control execution. The definition of the power motive offered here, however, is meant to include more than dominance. The importance of this kind of social motivation has been elaborated by Adler (1) and Sullivan (12) and has been brought to bear upon some of the recent research in group dynamics (6).

Before one can make these predictions about the relationship of power motivation to behavior, a means of measuring the strength of the motive has to be established. The design of this research to develop such a measuring instrument falls into two steps.

First, written stories are obtained in response to pictures from two groups, one for whom it can be assumed that power motivation is aroused, and the other for whom it can be assumed the concern about power is not specifically aroused. A scoring system isolating the kind of imagery presumably reflecting power motivation can then be applied to these two sets of stories. The stories from the aroused group should contain more imagery in the categories scored for power motivation than those from the nonaroused group. Differences in imagery obtained between the two groups provide an empirical clarification of how power motivation is expressed in thought processes.

Second, the expected differences in imagery produced in these two experimental groups are used as the basis for a measure of individual differences in power motivation. Scores so derived should predict individual differences in other behaviors related to satisfying power motivation.

EXPERIMENTAL DESIGN AND HYPOTHESES

Step 1: Group Differences in Power Motivation

For the *arousal condition* the experimenter was able to take advantage of a natural situation: the election of student leaders at the University of Michigan. The candidates must petition for their candidacy. Once having declared candidacy, they have a month's time

[1] The research reported in this paper was submitted to the faculty of the University of Michigan in partial fulfillment of the requirements for the Ph.D. degree. The author would like to acknowledge the assistance of Dr. Daniel Katz in directing this work. The project was carried out while the author was a Research Training Fellow of the Social Science Research Council. He is indebted to the Council for its financial aid.

for campaigning before the actual election. After two days of balloting, the candidates congregate at the polls to see what the balloting has decided. At least two hours elapse after they have congregated before any official returns are announced. It was therefore decided to employ this waiting period for the study. All candidates who had at least one other opponent were previously asked by mail to cooperate in a research project designed to get at important social variables. The only information given was that the experimenter was interested in them as campus leaders, and was interested in the effects of social situations upon imagination. Because the candidates were presumably extremely concerned about the control of the means of influence—being elected to office—the situation was thought to meet the requirements for arousal of the motive as it was defined.

Since not all the candidates participated, volunteering bias may have been introduced. Such bias, however, would seem irrelevant to the problem under investigation. All 34 male *Ss* who appeared were used. No differentiation within this group was made on the kind of office for which they were candidates, since all the contested offices involved the responsibility of making decisions and influencing others.

To heighten the relevance of the cues of the ballot-counting situation to the story-writing procedure, the candidates were asked to judge their chances of winning the election, on the following scale: 0/5, 1/5, 2/5, 3/5, 4/5, 5/5 chances. The whole procedure was done anonymously.

The *Ss* in the *nonaroused condition* were 34 male students in an undergraduate course in psychology at the University of Michigan. The procedure was introduced as one about which some graduate students were interested in obtaining some normative data. After the procedure, the *Ss* were asked to sign their names, in order that further measures could be correlated with their story-writing performance. It seemed reasonable to regard this group as one for whom the concern about control of the means of influence was not specifically engaged.

One may argue that differences in respects other than aroused power motivation underlie such differences between these two groups as may be obtained in their stories. Perhaps *Ss* in the aroused group have had peculiar learning experiences predisposing them to write power-related imagery for reasons having nothing to do with power motivation as defined. The validity of this argument is granted; there is no means of isolating this factor in the present design. Greater reliance is placed on the second step in the design—predicting power-motivated behavior from a measure derived from the first step—for establishing the tentative validity of the measure of power motivation. This problem will be more fully treated in discussion of the results.

Story-writing procedure. All *Ss* wrote stories for five pictures in a manner outlined in *The Achievement Motive* (8). The only procedural difference is that the pictures were reproduced in booklets rather than projected on a screen, and instructions modified accordingly. The five pictures are listed below in the order which they were presented:

1. Two men in a library; Picture E in the Achievement series (8).
2. Instructor in classroom; Picture 21 in Birney's study (5).
3. Group scene; Picture E in the Affiliation series (3).
4. Two men on campus; Picture 41 in Birney's study (5).
5. Political scene; from magazine, *Woman's Day.*

Step 2: Individual Differences in Power Motivation

Since the stories of the nonaroused *Ss* could be identified, other measures possibly related to a derived score of power motivation were obtained in this group: (*a*) Allport-Vernon Scale of Values (2), (*b*) Vocational Goals Inventory (7), (*c*) questionnaire on siblings, and (*d*) ratings by instructors of classroom behavior. The Vocational Goals Inventory required *Ss* to rank-order ten listed possible satisfactions on a job: being leader, interesting job, recognition from fellow men, being boss, surety of keeping job, self-expression, high pay, fame, helping others, working on one's own. The sibling questionnaire asked *Ss* to list their siblings and their ages, to permit reconstruction of birth order. The ratings by instructors pertained to the following classroom behavior: (*a*) Argumentative in the classroom—likes to argue and contradict; (*b*)

Eager to get points across in the classroom in order to convince either the instructor or other students of a point of view. They were made on a three-point scale of frequency: 1 = "very frequently," 2 = "at times," and 3 = "never or rarely." The following hypotheses were advanced concerning the relationship between individual differences in power-motivation scores derived from the projective instrument and the variables measured by the preceding measures:

Hypothesis 1. Power motivation is positively related to Political and Economic, and negatively to Social value orientation, as measured by the Allport-Vernon Scale of Values. A person highly interested in controlling the means of influence should value those ideas having to do with such control within the culture—namely, the political and economic values. Being interested in controlling a person in a sense negates his sympathy for this person's feelings, and therefore the negative relationship with the Social dimension was anticipated.

Hypothesis 2. Power motivation is positively related to the intensity of interests in the job satisfactions of being boss and being leader. These job satisfactions can be interpreted as positions in which the potentiality for controlling the means of influence is high.

Hypothesis 3. Power motivation is positively associated with the number of siblings in the *S*'s family. Furthermore an association was hypothesized between birth order and strength of power motivation, though no prediction about the direction of the association is made. Power satisfactions are perhaps derived from many experiences in which being manipulated by others and manipulating others were in evidence. One would expect that the high power-motivated individuals would come from larger families where this type of experience is more likely to occur. Birth order would seem to be an important variable in further shaping these experiences.

Hypothesis 4. Power motivation is positively associated with instructors' ratings of argumentativeness and trying to convince others. These behaviors are presumably instrumental in satisfying the power motive in the classroom.

SCORING STORIES FOR POWER MOTIVATION

The method of scoring the stories for power motivation closely parallels the pattern established for hunger (4), achievement (8), and affiliation (3) motives. The categories scored follow in a behavioral sequence analysis which is discussed with some detail in *The Achievement Motive* (8).

In scoring a story it is first identified as being related to power motivation or not. If a scorer decides that the story is related to power motivation (i.e., the story meets the Power Imagery criteria given below), it is then further scored for the presence or absence of the categories in the behavioral sequence. If the scorer decides that the story is not related to power motivation (Unrelated), no further scoring is done. The Power Imagery criteria are given in detail below. A complete scoring system with examples is available (13).

Power Imagery is scored when a story contains a reference to the thoughts, feelings, or actions of one of the characters which indicates that the character is concerned with the control of the means of influencing a person. Evidence of concern can come from one or more of three sources: (*a*) statements of affect; (*b*) statements of control activity; (*c*) statements of superior-subordinate role relations. These three criteria are elaborated below.

a. There is some statement of affect surrounding the maintenance or attainment of the control of the means of influencing a person. A character can be feeling good about winning an argument or feeling bad because he was unable to have his way. Also statements about wanting to win a point, or showing dominance, wanting to gain control (such as by a political or executive position), convince someone of something, or put a point across, can be interpreted as implicit statements of affective concern about the control of the means of influence. Affective concern can also be found in statements of wanting to avoid weakness. Examples of this are: being humiliated in a status position, being ashamed of an incapacity to assert one's self or become dominant, resenting the influence of another and wanting to overcome this.

Some very weak statements of concern over control of the means of influence are scored. Statements of desires to teach another person something, to interest another person in something, to inspire someone—although apparently weak in obvious power significance—should be scored. The only times when these statements would not be scored are in cases where the teaching, inspiring, or interesting is solicited by the person being influenced.

Statements of a person wanting another person to gain control of the means of influence—such as a person wanting someone else to win an election—should not be scored as Power Imagery. The only exception is a case in which there is a very clear identification of the person with the preferred candidate. This kind of imagery may appear in election stories where John wants to win the election, although John is not the candidate.

b. There is a definite statement of someone doing something about maintaining or attaining the control of the means of influence. Something that the character is actually doing in the story is the only kind of imagery that can qualify as Power Imagery under this criterion. The character has to dispute a position, argue something, demand or force something, try to put a point across, try to convince someone of something (and

theoretically any activity in order to obtain control of the means of influencing someone).

Three special considerations should be noted under Criterion *b*. The two special considerations noted under Criterion *a* are applicable to Criterion *b*. Statements involving trying to teach, interest, or inspire someone are scorable under Criterion *b*. Trying to win an election for someone else will not be scorable unless there is a close identification between the person campaigning and the candidate for office. Sometimes, moreover, it will be clear that an activity, although it meets the criterion listed in the preceding paragraph, is only for the purpose of arriving at the goal of another motive. In such a case, the story should not be scored for Power Imagery.

c. A story can be scored for Power Imagery if there is a statement of an interpersonal relationship which in its execution is culturally defined as one in which there is a superior person in control of the means of influencing a subordinate person. Examples of this are: boss-worker, judge-defendant. Mere mention of superior-subordinate relationship is not enough to score. There has to be some mention of the activity involved in carrying out this relationship. Mere mention of a given person being influential would be enough to allow this story to be scored for Power Imagery.

Unrelated Imagery is scored when there is no evidence of power motivation concern in the stories as defined under Power Imagery.

When the story has been identified as one containing Power Imagery, it is scored for the presence or absence of the subcategories similar to those outlined for achievement motivation (8): Need; Instrumental Activity; Goal Anticipation; Block; Affective State; Thema.

In the final coding the coder did not know whether a given story came from one or the other of the two groups. To check on the reliability of this coding, a second coder scored a random set of twenty-four *S*s from the total sample. The percentage of agreement for a given category between the two scorers ranged from 57 to 97. Furthermore, need for power scores (n Power scores) were tallied for each *S* by adding the number of categories appearing in his stories. The n Power scores obtained from the two coders were rank-ordered for the subsample. The rank-order correlation (rho) between the ranks obtained on these *S*s by the two judges was +.87.

TABLE 1

NUMBER OF SUBJECTS IN AROUSED AND NONAROUSED GROUPS ABOVE AND BELOW MEDIAN* FOR VARIOUS POWER-RELATED IMAGINATIVE CATEGORIES

Imaginative Category	Aroused Group (N = 34)		Nonaroused Group (N = 34)		p‡
	Above†	Below	Above	Below	
Power Imagery	27	7	15	19	.0023
	Breaking point between 1 and 2				
Need	21	13	14	20	.047
Instrumental Activity					
Positive	22	12	15	19	.047
Negative	12	21	8	26	.090
Questionable	16	18	13	21	n.s.
Any sign	24	10	13	21	.0056
	Breaking point between 1 and 2				
Goal Anticipation					
Positive	3	31	1	33	n.s.
Negative	3	31	3	31	n.s.
Any sign	6	28	3	31	n.s.
Block					
Personal	4	30	1	33	n.s.
Environmental	3	31	2	32	n.s.
Affective State					
Positive	4	30	1	33	n.s.
Negative	3	31	2	32	n.s.
Any sign	6	28	3	31	n.s.
Thema	28	6	17	17	.0044

* Median based on combined frequencies of both the aroused and the nonaroused groups.

† Except where indicated *S*s were divided as to whether the category appeared once or not at all in the five stories, this procedure approximating the combined median most closely.

‡ All the probabilities were computed by Exact Test, one-tailed hypothesis, 1 *df*.

RESULTS

Step 1: Group Differences in Power Motivation

Imagery in the stories written by the aroused and nonaroused groups was compared by dichotomizing the frequency distributions of the combined groups as near to the median as possible. Table 1 gives for each category the number of *S*s in each group above and below the median of the combined distribution. Exact tests were computed for each instance that potentially represented a significant deviation from a chance distribution. Power Imagery, Need, Instrumental Activity, and Thema appeared significantly more often in the aroused group's stories than in the nonaroused group's stories. These results, inci-

TABLE 2

NUMBER OF SUBJECTS IN AROUSED AND NONAROUSED GROUPS ABOVE AND BELOW MEDIAN* OF N POWER SCORES

Group	n Power Score	
	Above 6	Below 6
Aroused group (N = 34)	22	12
Nonaroused group (N = 34)	9	25

Note.—p = .0013, one-tailed hypothesis, 1 *df*.

* The approximation to the median was computed by combining the n Power scores of both the aroused and the nonaroused groups.

TABLE 3
FREQUENCIES OF HIGH-LOW n POWER SCORERS IN VARIOUS CATEGORIES ON MEASURES RELATED TO POWER MOTIVATION

Measure	Dimension	High n Power	Low n Power	Chi square	p	df
Allport-Vernon Scale of Values	Political					
	High*	8	5	—	n.s.	—
	Low	8	8			
	Economic					
	High	6	5	—	n.s.	—
	Low	10	8			
	Social					
	High	7	9	—	.12	1
	Low	9	4			
Vocational Goals Inventory	"A job where you could be leader"					
	High	9	4	—	.096†	1
	Low	7	10			
	"A job where you could be boss"					
	High	8	8	—	n.s.	—
	Low	8	6			
	"A job where you could be looked upon very highly by fellow men"					
	High	10	3	—	.049	1
	Low	6	11		(2 tailed hypothesis)‡	
Sibling Questionnaire	No. of Siblings					
	0	0	1			
	1	7	4	—	n.s.	—
	2	4	3			
	3 or more	5	3			
	Has older sibling					
	Yes	9	5	—	n.s.	—
	No	7	6			
Instructor's Ratings	Argumentativeness Frequency					
	High	7	1			
	Medium	7	7	5.62§	$<.05$	2
	Low	4	7			
	Frequency of Trying to Convince Others					
	High	9	1			
	Medium	7	6	11.02	$<.001$	2
	Low	2	8			

* High and Low division on Allport-Vernon Scale of Values based on deviation from normative data (2).
† The probabilities of the 2 × 2 tables were computed by the Exact Test.
‡ All other probabilities were estimated under a one-tailed hypothesis.
§ Chi squares where reported were estimated by the Mood Maximum Likelihood Ratio (9).

dentally, are comparable to those obtained in similar comparisons in respect to the achievement and affiliation motives, except that the categories Goal Anticipation, Block, and Affect failed to differentiate the groups.

An over-all check on the differences in imagery can be made by comparing the n Power scores for the *Ss* in the two groups. Table 2 shows the number of *Ss* in each group above and below the combined groups' median n Power score. The probability of obtaining this distribution by chance is .0013, again as estimated by an Exact Test.

Step 2: Individual Differences in Power Motivation

Scores of the nonaroused group for n Power were dichotomized as near as possible to the group median. The resulting division into High and Low n Power scores is retained through all comparisons to be made in this section. The results of the comparisons made between the n Power scores and the variables for which relationships were predicted appear in Table 3. In most cases the other variables are dichotomized near the median into high and low categories.

This table supports the following observations which confirm predicted hypotheses. High n Power scorers, as compared to low n Power scorers, tend to have lower scores on the Social Value Dimension of the Allport-Vernon Scale of Values; show a significantly stronger interest in the job satisfaction of being leader; were rated significantly higher by their instructors on the frequency of argumentation and the frequency of trying to convince others of their points of view in the classroom.

Other hypotheses were not confirmed. High n Power scorers as compared to low n Power scorers did not differ significantly in their scores on the value orientations Economic and Political, on their ratings of the job satisfaction of being boss, and on the number or order of siblings in their families.

Table 3 also includes an unpredicted result, that high n Power scorers rated the job satisfaction of obtaining recognition from their fellows significantly higher than low n Power scorers. No other comparison between the motive measure and the Vocational Goals Inventory yielded any positive results.

DISCUSSION

The results obtained comparing the imagery produced by the aroused and nonaroused groups are similar to parallel results found with the measures of achievement and affiliation motivations. In the present design however there is an ambiguity in the nature of the arousal condition which was not present in the earlier studies. This ambiguity—the fact that the power arousal group was different from the nonaroused group not only in presumed level of arousal of some motivational pattern but also in whatever dispositions are salient in discriminating candidates and noncandidates for college offices—forces one to make only tentative conclusions about these results. It may be that the group differences uncovered in the analysis of the stories are due to motivations which are distinct from the power motive as defined.

Two other reasonable interpretations of the obtained differences between the two groups attribute them to differences in the level of achievement or recognition motivation. Fortunately it was possible to check the former possibility. The two sets of stories were scored for achievement motivation as outlined in *The Achievement Motive* (8). No significant difference in mean n Achievement scores of the groups was found.

The other interpretation, that the group differences are accountable in terms of recognition rather than power motivation, is a distinct possibility. The only basis for thinking otherwise is that the scoring system was built around a conceptualization of power motivation as distinct from strivings for recognition. Recognition themes in stories were not scored unless they fulfilled the criteria for power motivation as well. That is, if a story brought up problems of recognition as a way of obtaining control of the means of influence, then and only then was it scored. There is, however, no experimental way in the present study to rule out the interpretation of the power-motive measure in terms of recognition. In further use of the measure of power motivation caution to recognize this possible contamination is recommended.

There seemed to be no difference between the aroused and nonaroused groups on certain of the affected categories (Goal Anticipation, Affect, and Obstacle), which proved valuable in discriminating groups with the other measures. An explanation for this discrepancy should be offered. Perhaps this particular motive does not permit the arousal of associated affective thoughts. There might be good cause to inhibit the expression of affect surrounding the power area, since there are many recognized taboos for expressing power in this culture—especially the expression of overt satisfactions derived from power behavior. Support for this explanation comes from the observation that these categories appeared very infrequently in the stories of both groups. Another reason for the lack of differential appearance of these categories might be that the particular pictures used in this study did not provide cues to the kind of thinking that would reflect the affective component. This possibility can be tested in further research.

The fact that the measure not only reflects the group differences in assumed motivation level, but also individual differences in variables related to aspects of power concern as defined, suggests that the instrument can be used with success in researches in which such a measure of motivation would be needed. The measure of power motivation, if some of its

ambiguities are clarified, can be potentially used for exploring dimensions of control behavior exhibited in interpersonal relationships. Researchers using concepts of power different from the one presented here, should be cautious in the use of this measure.

Some of the hypothesized relationships, particularly with the Allport-Vernon scales, were not substantiated in this study. These value measures can perhaps reflect many *different* motivational orientations, and it may thus be naive to expect a positive relationship between a fairly specific motivational pattern and values that can service many different needs.

Lack of support for the prediction that *Ss* high in n Power would rate the job satisfaction of being boss higher than *Ss* low in n Power may indicate that re-evaluation of the meaning of being boss for them is in order. "Boss" can be thought of as a negative word in this culture, and, indeed, there was little variability in the *Ss*' ranking of job satisfaction in this respect. Nearly everyone ranked this alternative low.

The simple prediction that high n Power *Ss* would come from large families was not substantiated. This finding would seem to indicate that any relationship between the power motive and the family structure depends on more complex considerations than simple frequency of contact with other family members. When the distribution of n Power score is divided into thirds (high, medium, and low), a trend was uncovered. Eight out of twelve high n Power scorers had at least one older sibling whereas this was true for only one out of six medium n Power scorers and five out of nine low n Power scorers. (Chi square = 4.34, $p < .15$, 2 *df*, two-tailed hypothesis.) Perhaps some fruitful distinctions can be made in the area of family structure which may be useful in understanding the development of the power motive.

The finding that high n Power *Ss* rated the job satisfaction of gaining recognition from their fellow men higher than do low n Power *Ss* points to the possibility that the satisfactions of prestige may be intimately tied with the satisfactions of power. This point of view has been elaborated by Murphy (9). There is also the possibility that the measure includes, at least partially, some aspects of recognition motivation. Indeed, all of the significant findings can be theoretically accounted for by considering the motive measured to be the recognition motive.

The relationship between achievement and power motivation. The product-moment correlation between the n Achievement and n power scores was $+.27$ ($p < .05$). An attempt was made to use the n Achievement scores, however, to predict the same variables with which n Power scores had been related. No significant findings with these variables were obtained. In short, there seems to be a slight positive association between the two motives, but what is measured in scoring for the power motive is that process which seems to be directly related to power, rather than achievement strivings.

SUMMARY

Two groups of *Ss* wrote stories in response to five pictures. One was tentatively considered to be aroused by cues relevant to power motivation; the other was not aroused. A scoring system paralleling the coding methods which had been fruitful in measuring the achievement and affiliation motives was applied to the stories, and succeeded in differentiating the two groups. Four of the seven categories coded appeared significantly more frequently in the stories of the aroused group.

Individual differences in power motivation were estimated for *Ss* in the nonaroused group by the n Power scores (based on the coding categories). High n Power scorers differed from low n Power scorers in this group on a number of other factors relevant to power motivation. Some hypotheses relating strength of power motivation to other variables were not confirmed.

Although there was a significant positive correlation between n Power scores and n Achievement scores, the n Achievement score distribution does not predict the same relationships with power-related variables.

In conclusion, the results indicate that the projective measure of power motivation developed here not only can successfully isolate the presumed differential level of motivation in groups, but also can successfully predict attitudes and overt behavior which are presumably related to the processes involved in power motivation. It is cautioned, however, that the measure may be conceived also in terms of recognition motivation.

REFERENCES

1. ADLER, A. *Understanding human nature.* New York: Greenberg, 1927.
2. ALLPORT, G. W., & VERNON, P. E. *Manual for the study of values* (Rev. Ed.) Boston: Houghton Mifflin, 1931.
3. ATKINSON, J. W., HEYNS, R. W., & VEROFF, J. The effect of experimental arousal of the affiliation motive on thematic apperception. *J. abnorm. soc. Psychol.*, 1954, **49**, 405–410.
4. ATKINSON, J. W., & MCCLELLAND, D. C. The projective expression of need. II. The effect of different intensities of the hunger drive on thematic apperception. *J. exp. Psychol.*, 1948, **38**, 643–658.
5. BIRNEY, R. Studies on the role of picture cues in projective measurement of achievement motivation. Unpublished doctor's dissertation, Univer. of Michigan, 1954.
6. CARTWRIGHT, D. Toward a social psychology of groups. Presidential address delivered before SPSSI, Cleveland, Ohio, 1953.
7. CENTERS, R. Motivational aspects of occupational stratification. *J. soc. Psychol.*, 1948, **28**, 187–217.
8. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R., & LOWELL, E. L. *The achievement motive.* New York, Appleton-Century-Crofts, 1953.
9. MOOD, A. M. *Introduction to the theory of statistics.* New York, McGraw-Hill, 1950.
10. MURPHY, G. Social motivation. In G. Lindzey (Ed.), *Handbook of social psychology.* Cambridge, Mass.: Addison-Wesley, 1954. Pp. 601–633.
11. SHIPLEY, T. E., JR., & VEROFF, J. A projective measure of need for affiliation. *J. exp. Psychol.*, 1952, **43**, 349–356.
12. SULLIVAN, H. S. *Conceptions of modern psychiatry.* Washington, D. C.: William Alanson White Psychiat. Found., 1947.
13. VEROFF, J. Development and validation of a projective measure of power motivation. Unpublished doctor's dissertation, Univer. of Michigan, 1955.

Received February 27, 1956.

GSR REACTIVITY AS A FUNCTION OF ANXIETY, INSTRUCTIONS, AND SEX

JAMES L. BERRY AND BARCLAY MARTIN

University of Wisconsin

SEVERAL studies indicate that GSR conditioning is positively related to the level of anxiety of the subject. Bitterman and Holtzman (4) found such a relationship when an over-all index of anxiety was assessed on the basis of several behavioral ratings and test scores. Welch and Kubis (10) found that psychiatric patients diagnosed as anxious showed more rapid GSR conditioning than a group of normals. Beam (3) found GSR reactivity to increase immediately before anxiety-arousing situations, such as oral examinations, as compared to a measure taken at another time. However, results have been rather uniformly negative in terms of a relationship between anxiety and GSR conditioning when the Taylor Anxiety Scale has been the sole measure of anxiety level. Neither Bitterman and Holtzman nor Beam found any relation between GSR conditioning and the Taylor Anxiety Scale.

The present study was designed to test the hypotheses that the Sarason Test Anxiety Scale and the kind of instructions given to the *S*s are related to GSR conditioning. The Test Anxiety Scale developed by Sarason (6) was tried in this study because of the previous failures of the Taylor Scale to be predictive in this area, and because it was thought on an a priori basis that the nature of the Sarason Test was such as to give a more valid measure of anxiety proneness among college students. The Test Anxiety Scale requires *S* to indicate on a series of five-point scales how he feels right before and during examinations in college courses, group intelligence tests, and individual intelligence tests.

Three different kinds of instructions were given with the intent of varying experimentally the level of apprehension that the *S*s would have about the conditioning situation. Taylor (9) varied instructions in a conditioned eyeblink study by telling one group that the air puff would increase in intensity as the trials progressed and telling another group that it would decrease. She found no differences between the groups. No previous attempts to modify GSR conditioning by instructions were discovered.

METHOD

From a class of 330 introductory psychology students, extreme groups of 60 high-anxiety (30 male and 30 female) and 60 low-anxiety (30 male and 30 female) *S*s were selected. The *S*s from each of these extreme groups of the same sex were then placed randomly in one of the three instructions groups. This procedure provided an anxiety-by-instructions-by-sex factorial design with 10 *S*s in each of the twelve groups.

A modified form of the Sarason Test Anxiety Scale,[1] with items referring to individual intelligence tests omitted, had been administered to the class as a group. Answers were recorded on regular five-point IBM answers sheets. The *S*s were then seen individually for the second session.

The Apparatus used to measure GSRs was an AC circuit designed by Grant (5). Dry silver finger electrodes were held on the palmar surfaces of two adjacent fingers on the right hand by rubber straps. Shock was supplied by a vacuum tube circuit which provided a constant current of .30 ma. to the left arm of the *S*. An audio-oscillator provided a 4,000 cycle tone. Cams attached to a one-cycle-per-minute motor were constructed so that a tone of 5 seconds' duration followed immediately by a shock of 0.5 second's duration could be presented every 30 seconds. The shock could be eliminated when desired by a switch.

The instructions given to the three different groups were as follows:

Group 1, Apprehension Arousal. Throughout the course of this experiment you will receive several electric shocks, such as the following (2 sample shocks). At one point in the experiment you will receive a shock considerably stronger, approximately twice as strong as the ones you just experienced. We realize that the shock is already strong and, because increasing it too much might have bad effects, you will receive only one of the stronger ones. This one may occur at any time during the course of the experiment. Do you have any questions before we begin?

Group 2, Neutral. Throughout the course of this experiment you will receive several electric shocks, such as the following (2 sample shocks). Do you have any questions before we begin?

Group 3, Reassurance. Throughout the course of this experiment you will receive several electric shocks, such as the following (2 sample shocks). These may seem rather strong to you now but you will find that you will adapt to them rather quickly and, after a few more, it will begin to feel like a pricking or

[1] We are grateful to Dr. Sarason for supplying us with a copy of the Test Anxiety Scale.

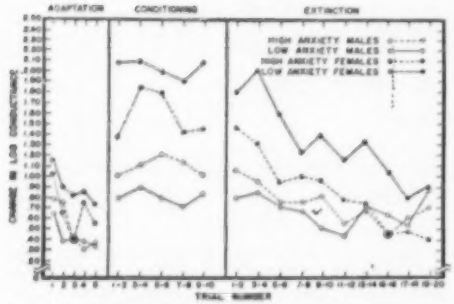


FIG. 1. ADAPTATION, CONDITIONING, AND EXTINCTION OF MALE AND FEMALE, HIGH- AND LOW-ANXIETY GROUPS

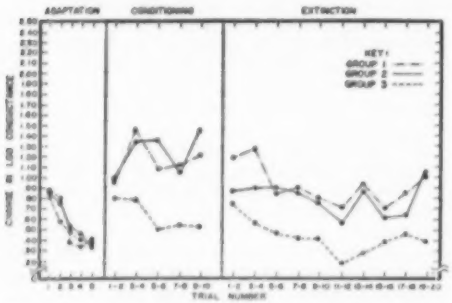


FIG. 2. ADAPTATION, CONDITIONING, AND EXTINCTION OF MALE INSTRUCTION GROUPS 1, 2, AND 3

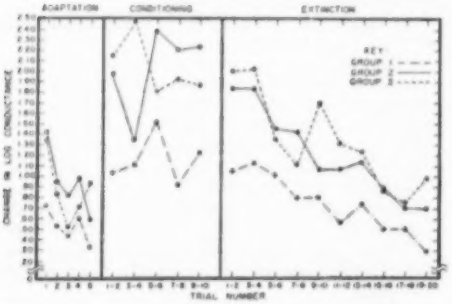


FIG. 3. ADAPTATION, CONDITIONING, AND EXTINCTION OF FEMALE INSTRUCTION GROUPS 1, 2, AND 3

tickling sensation. Most subjects haven't been bothered by them. Do you have any questions before we begin?

Each *S* was then given identical treatment in the conditioning situation. Four minutes of relaxation were followed by five trials with the tone alone, after which 10 conditioning trials and 20 subsequent extinction trials were administered. A partial reinforcement schedule was used during conditioning in which no shock was given on trials, 3, 5, and 9. Skin resistance readings were taken immediately before the tone and of the lowest point reached during the tone for all trials.

TABLE 1
SUMMARY OF THE ANALYSIS OF VARIANCE OF CONDITIONING TRIALS

Source	*df*	Mean Square	*F*	*p*
Total	119			
Within Groups	108	7.89		
Sex	1	103.9	13.2	.01
Anxiety	1	.29	.04	
Instructions	2	14.8	.93	
Sex × Anxiety	1	23.81	3.01	.10
Sex × Instructions	2	26.85	3.40	.05
Anxiety × Instructions	2	4.44	.56	
Sex × Anxiety × Instructions	2	1.92	.37	

RESULTS

The curves for adaptation, conditioning, and extinction are shown in Figs. 1–3. Resistance was transformed to log conductance − 1 to minimize the correlation between GSR and base level that is found when resistance is used as the measure, and trials were combined in pairs for conditioning and extinction curves. A summary of the analysis of variance for repeated measures for the average response during the conditioning trials is shown in Table 1. Bartlett's test for homogeneity of variance was nonsignificant. The *F* values for sex and the interaction of sex and instructions were significant at the .01 and .05 levels respectively. The interaction of sex and anxiety did not reach statistical significance.

In order to investigate the possibility of consistent trends as a function of trials, an Alexander trend analysis (1) was performed. The over-all group slope was not significantly different from zero, nor were the subgroups significantly different from each other in slope. It is apparent from Fig. 1 that most of the increased reactivity during the conditioning trials as compared with the adaptation trials occurred during Trials 1 and 2.

The extinction data were analyzed by computing the average response of each *S* over the 20 extinction trials, and then performing an analysis of variance upon these scores. Table 2 summarizes this analysis, and indicates that only the between-sex variance was significant.

[3] In order to make the figures easier to read, the transformation was actually as follows: 100 (log conductance −1).

TABLE 2

SUMMARY OF THE ANALYSIS OF VARIANCE OF EXTINCTION TRIALS

Source	*df*	Mean Square	*F*	*p*
Total	119			
Within groups	108	99.1		
Sex	1	502.7	5.07	.05
Anxiety	1	150.5	1.52	
Instructions	2	31.4	.32	
Sex × Anxiety	1	254.6	2.57	
Sex × Instructions	2	253.7	2.56	
Anxiety × Instructions	2	57.4	.58	
Sex × Anxiety × Instructions	2	16.6	.16	

The sex difference was found to hold during adaptation also. The average of Trials 4 and 5 of tone alone was found to be significantly higher for the female group ($F = 14.1$, $p < .001$).

The results are more clearly visualized in the curves. A comparison of the male and female curves show a marked tendency for females to have greater GSRs during all phases of the experiment. The interaction of sex by instructions is apparent when Fig. 2 and 3 are compared. For the males, the Reassurance instruction group is considerably below both of the other instruction groups. For the females, the Apprehension Arousal group is the lowest of the three. Thus, there is a tendency for the instructions to have opposite effects on males and females. This tendency was further analyzed within each sex group, using the average response during conditioning as the measure. The between instructions variance was significant for the female groups ($F = 3.59$, $p < .05$). The between-instructions variance for the male groups, however, was not significant ($F = 2.44$, $p < .10$). It is apparent from the curves that the Apprehension Arousal and Neutral male groups are very similar in their conditionability and that the Reassurance group falls below these groups. Since it was expected that the average response during conditioning would be less for the Reassurance group, it was considered appropriate to compute the probability of this occurrence. This was done by the use of Whitney's extension of the *U* statistic as described by Mosteller and Bush (7). The probability of the Reassurance group's being lower than both the other groups by chance alone was less than .01.

Although the term "conditioning" has been used throughout this study, it is probable that much of the variance may be the result of individual differences in general GSR reactivity, therefore reflecting pseudo-conditioning to some extent. That this is probably the case is suggested by other GSR data recently obtained in the Wisconsin laboratory on an *N* of 80 *Ss* in which the average GSR reactivity during trials with tone alone correlated .74 with average response during conditioning and .64 with average response during extinction. The possibility that pseudo-conditioning is involved of course does not make possible relationships between GSR reactivity and anxiety or instructions any less meaningful.

The marked sex differences in GSR conditioning found in this study are consistent with a similar finding with eye-blink conditioning reported by Spence and Farber (8). Females gave more conditioned responses ($p < .10$) than males. Similarly, Aronfreed, Messick, and Diggory (2) found greater GSR reactivity in females than males. Whether females are constitutionally more reactive in GSRs or more apprehensive in situations of this sort obviously cannot be determined from these results.

The findings seem to support the hypothesis that instructions affect the extent of GSR conditioning. Reassuring instructions to the male *Ss* resulted in lessened conditionability. The significant interaction of sex and instructions is puzzling, and suggests that whereas the male groups are affected in the predicted and common-sense direction by instructions, the female groups are affected oppositely, Apprehension Arousal instructions producing a decrease in conditionability. A similarly puzzling tendency, although not significant, was for the female high-anxiety group to show less conditioning than the low-anxiety group. It is perhaps best to withold speculation about this contrariness of females until such an unexpected finding can be cross validated.

The Sarason Test Anxiety Scale did not prove to be significantly related to GSR conditioning. Along with the previous failures of the Taylor Anxiety Scale to be related to GSR conditioning, this finding raises the question of whether any set of paper-and-pencil items is related to GSR conditioning. This question is being presently pursued by empirical item analysis.

SUMMARY

GSR conditioning and extinction measures were obtained on 120 college *Ss* in a factorial experimental design in which twelve treatment groups were divided into two sex groups, three instruction groups, and high- and low-anxiety groups. The three instruction groups were given three different kinds of instructions designed to be (*a*) apprehension arousing, (*b*) neutral, and (*c*) reassuring. The high- and low-anxiety groups were selected as extreme groups on the basis of the Sarason Test Anxiety Scale.

Results indicated that the instructions had opposite effects on the conditionability of males and females. The male *Ss* tended to show less conditioning when given reassuring instructions than when given either neutral or apprehension arousing instructions. The female *Ss* tended to show the least conditioning with the apprehension-arousing instructions and most conditioning with the reassuring instructions.

There was no relationship between GSR conditioning and the Sarason Test Anxiety Scale. Females showed greater GSR reactivity than males in all phases of the experiment; adaptation, conditioning, and extinction.

REFERENCES

1. ALEXANDER, H. W. A general test for trend. *Psychol. Bull.*, 1946, **43**, 533–557.
2. ARONFREED, J. M., MESSICK, S. A., & DIGGORY, J. C. Re-examining emotionality and perceptual defense. *J. Pers.*, 1953, **21**, 517–528.
3. BEAM, J. C. Serial learning and conditioning under real-life stress. *J. abnorm. soc. Psychol.*, 1955, **51**, 543–551.
4. BITTERMAN, M., & HOLTZMAN, W. Conditioning and extinction of the galvanic skin response as a function of anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 615–623.
5. GRANT, D. A. A convenient alternating current circuit for measuring GSR's. *Amer. J. Psychol.*, 1946, **59**, 149–151.
6. MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166–173.
7. MOSTELLER, F., & BUSH, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge, Mass.: Addison-Wesley, 1954. Pp. 317–319.
8. SPENCE, K. W., & FARBER, I. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 116–119.
9. TAYLOR, JANET A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.
10. WELCH, L., & KUBIS, J. The effect of anxiety on the conditioning rate and stability of the PGR. *J. Psychol.*, 1947, **23**, 83–91.

Received February 2, 1956.

THE RELATIONSHIPS AMONG IMAGINATIVE, DIRECT VERBAL, AND PHYSIOLOGICAL MEASURES OF ANXIETY IN AN ACHIEVEMENT SITUATION[1]

ALFRED C. RAPHELSON[2]

Bureau of Educational Research, University of Illinois

SEVERAL independent attempts have been made to relate different measures of anxiety (or fear) to performance in competitive achievement situations. In one series of studies (3, 11, 12, 15), anxiety level has been assessed by the Manifest Anxiety Scale (MAS), composed of refined MMPI items which have been classified by clinical psychologists as being indicators of general manifest anxiety. The experimental results indicate that subjects who score high in manifest anxiety condition faster in a classical type of experiment than low-scoring subjects, but show poorer performance in more complicated learning situations.

Mandler and Sarason (7) have developed a Test Anxiety Questionnaire (TAQ) which indicates the extent to which a subject experiences anxiety before, during, and after three typical kinds of testing situations, thus having more specific situational reference than the MAS. Using the TAQ, Sarason, Mandler and Craighill (8) obtained results indicating that as the expectation of failure is increased, the high-anxiety subjects show a decrement in performance, whereas the low-anxiety subjects show an increment.

McClelland *et al.* (6) placed subjects in situations differing in aroused motivation to achieve and asked them afterward to write stories to TAT-like pictures. From these stories a scoring system was developed to capture the thought content reflecting the degree to which a subject was motivated to excel in an achievement situation. This need-Achievement (n Ach) score related to behavior in a way that suggests that qualitative differences in thought content of persons who score high and low in n Ach represent real differences in expectancies concerning competitive situations (1). McClelland (5) speaks of the high scorers as showing a positive hope of success. He infers a defensive fear of failure from the relative absence of achievement imagery and relative preponderance of negative categories in the stories of the low scorers.

The present study attempts to investigate the interrelationships among these three measures and to test their relative effectiveness in accounting for reactions at another level of observation, i.e., physiological activity. All three measures give rise to distributions of scores described in terms of the amount of fear or anxiety that characterizes an individual generally (MAS) or specifically in competitive achievement situations (TAQ, n Ach). Traditionally, increases in skin conductance have been found to be correlated with emotional states. It is expected, therefore, that groups characterized as anxious on the individual difference measures (high MAS, high TAQ, low n Ach), if validly classified, will show conductance changes that are more consistent with the patterns expected in an emotional state than the groups classified as nonanxious (low MAS, low TAQ, high n Ach).

METHOD

Apparatus

The data were obtained in a room, 9 × 9 × 10 feet, in which was located the complex structure housing the essential apparatus. These essentials consisted of (*a*) a ground-glass screen 24 inches in diameter on which a ¾-inch spot of light moved, (*b*) a device which provided for a very irregular pattern of movement of this spot, (*c*) a recording polygraph, (*d*) a galvanometer, (*e*) a pneumograph, and (*f*) a red signal light.

The movement of the spot was accomplished by the projection of a ¾-inch spot on the screen by a mirror arrangement which was driven by two motor-driven cams in a highly irregular manner. A hand lever was connected in opposition to this device. A larger circle (3 inches in diameter) was also projected on the screen.

The recording galvanometer was used in conjunction with an electric network which automatically kept constant the current passing through the skin of *S* at 40 microamperes and measured the apparent resistance in terms of the potential difference across the electrodes. A continuous record of the changes in resistance of the skin was maintained by connecting the galvanom-

[1] This article is a portion of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy to the faculty of the University of Michigan. The writer wishes to express his deep gratitude to Drs. John W. Atkinson and Carl R. Brown for their patient direction and cooperation.

[2] Now at University of Michigan, Flint College.

eter to a stylus on the polygraph. One electrode was attached to the palm of the left hand, the other to the underside of the wrist on the same hand. Each electrode was made of a fritted glass filter of fine porosity, filled with physiological salt solution in which a loose coil of anodized silver wire was emersed.

The pneumograph was strapped to the abdominal region of *S* and connected to the polygraph.

A red signal light, located above the screen, indicated when *S* had allowed an error to occur. The errors were also indicated on the polygraph.

Subjects

The *S*s were 25 volunteer male students enrolled in various schools in the University of Michigan. Their class status ranged from freshman to graduate, and their age from 18 to 25.

General Plan

The general plan was to establish a testlike situation in which *S* was asked to keep the moving spot of light within a small area by manipulating the hand lever. In a brief introduction, the apparatus was described as being like devices used by the Air Force to test intellectual and motor abilities. The following test instructions were given:

> I am going to give you a series of reasoning problems that I want you to solve as best you can. I am not going to ask you for a solution; I do not care whether or not you can verbalize it. But I want you to show me you know the answer by your responses. Each problem consists of a sequence of movements of this spot of light over the screen. I will start the machine which controls the movements of the spot and your task is to discover what the sequence of movements is that the light follows. Now the lever-stick on your right works in opposition to the machine. If the machine moves the spot up you can counter the movement by moving the lever so that the spot moves in the opposite direction. Therefore, if you are able to solve the problem correctly, you will be able to anticipate each subsequent movement of the light, and by manipulating the lever correctly, you can keep the spot from moving very far. In the center of the screen you can see a ring outlined. Your task is to keep the spot of light from wandering out of the limits of this ring. It is an error if you allow the spot to leave that area. Your task is to make a good score by making as few errors as you can. The errors will be recorded on my tape here and after the session we will compare your score to that of others who have been through this test. You are to try your best to make as few errors as possible.
>
> Now in this problem you will have several other cues to aid you. First, the ring will always be there to indicate where the safe-area ends and the error-area begins. Second, every time you make an error this red light above the screen will flash on. It not only will indicate that you are in error but will also aid you in keeping a running count as to how well you are doing. Before you begin the test, I am going to begin the sequence, and I want you just to observe it for awhile so you can see what the problem is like and try to figure out the pattern of movements. I will tell you when I will begin the sequence and when you are to begin avoiding making errors. Do you have any questions?

Continuous records were taken of skin conductance and respiration during the entire experimental session.[3] The session consisted of the following four periods: *Null 1 Period*—a five-minute period in which *S* actively manipulated the lever but was not yet given the test instructions. *Anticipation Period*—a two-minute period which immediately followed the test instruction. The *S* watched the movements of the spot but did not yet begin to avoid making errors. *Solution Period*—a ten-minute period in which *S* actively attempted to make as few errors as he could by controlling the movements of the spot. *Null 2 Period*—a five-minute period after the Solution Period which was similar to the Null 1 Period. No problem was presented while the *S* simply manipulated the lever. The first and last periods were conceived of as relatively free of anxiety-arousing cues. It was not expected that the subgroups differentiated by the various dispositional measures would be different in their physiological reactions in these periods.

Since the spot actually moved in an irregular manner and no solution was possible, errors are not a meaningful variable in this experiment. An attempt was made to keep the frequency of errors as constant as possible for each *S* by manipulating the amplitude of the spot's movement in relation to *S*'s actual performance.

Measures of Individual Difference

Following the motor task, *S*s wrote the n-Ach stories while still connected to the recording equipment. They were disengaged before taking the other two tests. To be consistent with the literature on these measures, each distribution of scores was broken into high, middle, and low thirds for comparison on the physiological indices. The n-Ach measure consisted of four TAT-like pictures, presented to elicit imaginative responses from which a score was determined according to the coding procedure described by McClelland, *et al.* (6). The experimenter and another trained coder scored all the stories. The reliability coefficient between the two sets of scores was .82. Differences were resolved by reference to a third expert.[4] The scores for the high-n-Ach third ranged from 8–14; for the low third, −3–2.

The Test Anxiety Questionnaire has been fully described by Mandler and Sarason (7). A high score indicates a tendency to reflect feelings of anxiety in a testing situation. The scores for the high-TAQ third ranged from 23–31; for the low-TAQ third, 1–13.

The Manifest Anxiety Scale has been described by Taylor (14). The scores for the high-MAS third ranged from 15–22; for the low third, 1–7.

Skin Conductance Measure

The skin resistance records were analyzed in the following way: Samples from the records were taken

[3] No significant relationships were found between any of the other variables in these experiments and any of the characteristics of respiration activity (amplitude, rate, volume). Hence no respiratory data will be reported.

[4] The writer is indebted to Mr. Berne Jacobs and Dr. Joseph Veroff for their cooperation in this scoring.

at the beginning of each period and at the end of each subsequent minute. These readings were converted into units of conductance (micromho) and an average based on these samples was computed to represent the entire period. Analyses were done on the average level in each period in terms of variation in conductance within each period expressed as a percentage change over the Null 1 Period average. The percentages were calculated as follows:

$$\frac{\text{Mean Conductance in period}}{\text{Mean Conductance in Null 1}} \times 100.$$

These percentages were converted into angles by the arc sin $\sqrt{\text{percentage}}$ transformation to insure the independence of means and variances (10).

Statistical Analysis

The error estimates used to evaluate the mean differences were derived from analyses of variance designated Type III by Lindquist (4). Since there were twenty-five *Ss* in this study, it was impossible to have equal numbers in each of the three subgroups for each measure. The numbers in each subgroup, therefore, were equalized by randomly eliminating one case from the subgroup that contained the unequal number.

RESULTS

Interrelationships Among Measures of Individual Differences

The product-moment correlations among the three measures of individual differences are given in Table 1. The two measures of reactions specific to achievement (n Ach, TAQ) were significantly correlated, as were the two measures concerned with negative reactions towards general situations (MAS, TAQ). This pattern of relationships is consistent with the specifications set for the measures by the theories to which they are related. The TAQ was developed to measure anxiety symptoms specific to an achievement situation. The MAS was developed to measure general feelings of anxiety which may or may not occur in an achievement situation. A significant positive correlation is expected between them. The n-Ach measure is thought to reflect the degree to which *S* has expectancies of goal attainment in an achievement situation; such expectancies are incompatible with any anxiety reactions associated with the setting. A high negative correlation is expected with TAQ and a low negative correlation with MAS.

TABLE 1

PRODUCT-MOMENT CORRELATIONS AMONG THREE MEASURES OF INDIVIDUAL DIFFERENCES

Measure	Test Anxiety	*N*	Manifest Anxiety	*N*
n Achievement	−.43*	24†	−.25	25
Test Anxiety			.53	24

* *r*'s of .39 and .50 are significant at .05 and .01 levels of confidence respectively.

† One subject failed to fill out the Test Anxiety Questionnaire correctly and was eliminated from all Test-Anxiety analyses.

TABLE 2

MEAN ABSOLUTE CONDUCTANCE IN NULL 1 PERIOD

Measure	High (*n* = 8)	Middle (*n* = 8)	Low (*n* = 8)	σ diff.
n Achievement	23.9*	20.7	17.5	6.42
Test Anxiety Questionnaire	17.3	22.2	23.5	6.40
Manifest Anxiety Scale	17.7	22.4	22.0	6.35

* Units in micromhos.

TABLE 3

MEAN VARIATION IN CONDUCTANCE EXPRESSED AS PERCENTAGE OF NULL 1 LEVEL (*N* = 8 for all subgroups)

Measure	Subgroup	Period		
		Anticipation	Solution	Null 2
n Achievement	High	102.5*	95.4	68.9
	Middle	101.1	106.5	85.0
	Low	100.1	110.0	79.4
Test Anxiety Questionnaire	High	104.2	109.2	79.6
	Middle	101.0	103.1	86.8
	Low	103.8	89.5	66.8
Manifest Anxiety Scale	High	106.0	109.6	82.8
	Middle	94.0	98.8	71.2
	Low	103.2	101.8	78.6

* The σ diff. for each measure is as follows: n Ach: between = 16.4, within = 7.4; TAQ: between = 16.3, within = 6.4; MAS: between = 16.5, within = 7.6.

Absolute Conductance Level in Null 1 Period

The initial Null 1 levels of conductance in relation to the several measures of individual differences are presented in Table 2. In this period, the anxiety-arousing instructions had not yet been introduced. However, on each of the three measures, the subgroup which is presumed to be composed of relatively nonanxious *Ss* showed the highest conductance level. Although even the largest difference, that obtained between the subgroups scoring high and low on n Ach, was not significant (σ diff. = 6.42, t = .997), the consistency of the trends in each of the comparisons may be worth noting. They

TABLE 4

MEAN VARIATION IN CONDUCTANCE EXPRESSED AS A PERCENTAGE CHANGE OF NULL 1 LEVEL FOR SUBJECTS DESIGNATED ANXIOUS OR NONANXIOUS*

Subgroup	*N*	Period		
		Anticipation	Solution	Null 2
Anxious	7	104.4	112.1	83.7
Nonanxious	9	101.7	89.0	67.6
Difference		2.7	23.1	16.1
σ diff.		12.42	12.22	9.05
t		.22	1.89	1.78
p (one-tailed)		—	.04	.05

* On the basis of scores on both the Test Anxiety and the n achievement measures.

are consistent with the notion that a high conductance level may be indicative of psychological states other than anxiety.

Skin Conductance Variation in Experimental Periods

The mean variation in conductance, expressed as a percentage change over Null 1 level in three experimental periods, is presented in Table 3 for each of the subgroups differentiated by the three measures. The subgroups differentiated by measures of n Ach and TAQ showed significantly different conductance patterns. Though none of the between-subgroup differences were significant at any particular period, there were meaningful differences between subgroups in the direction of the trend of skin conductance through the different experimental periods. The subgroups considered relatively "anxious" (low n Ach, high TAQ) showed an increase in conductance from the Anticipation to the Solution Period. The subgroups considered relatively nonanxious (high n Ach, low TAQ) showed a decrease. The differences between these particular differences with both measures were significant in the predicted direction (n Ach: diff. = 17.0, σ diff. = 10.5, t = 1.62, p = .05; TAQ: diff. = 19.3, σ diff. = 9.1, t = 2.12, p = .025).

The subgroup designated as anxious by the more general MAS did not show any pattern of change in conductance that was significantly different from the subgroup designated as nonanxious.

A final analysis was made, comparing only those *S*s scoring both high in TAQ and low in n Ach (both treated in the past as symptomatic of a disposition to be anxious in achievement situations) with *S*s who scored both low in TAQ and high in n Ach (both treated in the past as symptomatic of the relative lack of a disposition to be anxious in an achievement situation). The significant negative correlation between the two measures implied that by examining *S*'s score on both measures, it would be possible to obtain a more reliable estimate of the degree to which he might be anxious or nonanxious in a competitive situation. On each measure, the distribution was divided as near as possible to the median. Only *S*s who fell in the corresponding high and low anxiety halves of both distributions were selected for comparison. Mean variations in conductance for these subgroups are presented in Table 4. It can be seen that by this method of classifying *S*s, the trends observed between subgroups in the separate analyses were further emphasized and were significant in the Solution and Null 2 Periods. The directional change between the Anticipation and Solution Periods observed, significantly different in the separate analyses, was also significant here (diff. = 20.4, σ diff. = 10.1, t = 2.01, p = .03).

The trends observed in the earlier analyses concerning the initial level of conductance in Null 1 were further emphasized in this method of classification. The subgroup designated nonanxious showed a significantly higher initial level of conductance than the subgroup designated anxious (diff. = 7.2, σ diff. = 2.88, t = 2.50, p = .02).

DISCUSSION

Skin conductance, having been traditionally considered a physiological index of anxiety, is here viewed as a criterion against which discriminations made by the three dispositional indices of anxiety may be checked. The findings imply that the more specific an assessment technique is to the situation in which the predicted behavior is to occur, the more accurate is the prediction. This point is emphasized by the results obtained when *S*s were classified according to their standing on both the n-Ach and the TAQ measures. By this method, the nonsignificant between-subgroup trends observed in the individual analyses were further emphasized, reaching a significant level in two of the three experimental periods. Further, the difference in the amount of change between the two subgroups from the Anticipation to

the Solution Periods was accentuated as predicted.

An important question is raised by the present result: Of what importance are measures which are designated to be "general indicators" of anxiety? Behavior always occurs in specific situations. If an individual is termed "generally anxious," does it imply that he is anxious in all specific kinds of situations or just in a greater than average number? These results suggest that particular anxiety reactions are associated with specific situations. An assessment technique constructed to reflect response tendencies peculiar to a certain limited class of situations will probably have a predictive advantage over more generally oriented methods.

The analyses made in terms of the initial level of conductance in the Null 1 Period showed that in each of the individual difference measures, the subgroup considered to be the least anxious generally had the highest level of absolute conductance. Though these differences between subgroups in the individual analyses were not significant, they did reach statistical significance when *Ss* were classified on the combined n-Ach-TAQ indices—presumably the most reliable estimate of individual differences in disposition for achievement test anxiety. Since the Null 1 Period occurred before the experimental instructions were given, the differences must be due to initial reactions to the experimental situation itself, unless differences in conductance are a stable correlate of the individual difference measures.

Schlosberg (9) has argued that conductance levels should be considered an indicator of the degree of energy mobilization or "activation" of the individual in a particular situation. Activation has no particular "sign" or direction; it may stand either for positive interest in the task or for fear motivating avoidance. The *Ss* with a higher conductance level on entering the experimental room may be reacting to the situation with positive anticipatory interest cued off by the complex apparatus that immediately faces them. This interpretation is consistent with the contention of McClelland *et al.* (6) that the high n-Ach subgroup is likely to be challenged by a problem situation. Burdick (2) similarly reported a significantly greater tendency for persons high in n Ach to volunteer to participate in psychological experiments.

If Schlosberg's theory of activation is correct, it may mean that the increased conductance level that the high n-Ach subgroup shows initially has quite a different "sign" than the change the low n-Ach subgroup shows later. The conductance measure itself is not capable of giving the direction of behavior implied by increased activation, but it is possible to infer direction from the way the two subgroups are conceptualized according to n-Ach theory. It is possible, therefore, to look at the conductance measure as an indicator of the degree to which an individual is activated by the cues of the situation. The direction of this activation, either positive (approach) or negative (avoidance), can be inferred through an analysis of the content of human thought (6, p. 33). Neither the MAS nor TAQ allow this possibility since there is no direct way of obtaining evidence of positive orientation from them.

SUMMARY

This study investigated relationships among three dispositional measures and two physiological indices of anxiety in a competitive achievement situation. The dispositional measures were relatively weak motivation to achieve as measured through content analysis of imaginative stories (low need Achievement), the Mandler-Sarason Test Anxiety Questionnaire, and the Taylor Manifest Anxiety Scale. The physiological indices were skin conductance and respiratory volume.

Twenty-five male college students were placed in a competitive test situation and instructed to perform a complex perceptual-motor task. They were informed when they made errors and were given cues to aid in performing the task accurately. The physiological indices were recorded before the ego-involving instructions were given and immediately before, during, and after the performance task. The three dispositional measures were obtained after the performance task.

The correlation between need Achievement and Test Anxiety was −.43; between Test Anxiety and Manifest Anxiety, .53; between Achievement and Manifest Anxiety, −.25. The two measures specifically concerned with reactions in competitive achievement situations (n Achievement, Test Anxiety), were both related to changes in skin conductance

during the performance task. The relationship was clearest when the subjects were classified on both of these measures as anxious (i.e., high Test Anxiety and low n Achievement) or nonanxious (low Test Anxiety and high n Achievement). The more anxious group increased in conductance while the relatively nonanxious group decreased. The Manifest Anxiety Scale did not relate to conductance change and no consistent relationships were found between respiratory activity and any of the other measures. These results suggest the advantage of situational over general measures of anxiety in accounting for changes in skin conductance in a particular situation.

It was also found that the highest initial level of conductance was obtained by the subgroup designated as nonanxious on each of the three dispositional measures respectively. These differences occurred before the introduction of the anxiety-arousing instructions. These findings are interpreted in terms of Schlosberg's theory of activation level.

REFERENCES

1. ATKINSON, J. W. Exploration using imaginative thought to assess the strength of human motives. In M. R. Jones (Ed.), *Nebraska symposium on motivation*. Lincoln: Univer. of Nebraska Press, 1954.
2. BURDICK, H. A. The relationship of attraction, need Achievement and certainty to conformity under conditions of simulated group atmosphere. Unpublished doctor's dissertation, Univer. of Michigan, 1956.
3. FARBER, I. E., & SPENCE, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 120–125.
4. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
5. MCCLELLAND, D. C. *Personality*. New York: Sloane, 1951.
6. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & Lowell, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
7. MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166–173.
8. SARASON, S. B., MANDLER, G., & CRAIGHILL, P. C. The effect of differential instructions on anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 561–565.
9. SCHLOSBERG, H. Three dimensions of emotion. *Psychol. Rev.*, 1954, **61**, 81–88.
10. SNEDECOR, G. W. *Statistical methods*. Ames: Iowa State Press, 1946.
11. SPENCE, K. W., & FARBER, I. E. Conditioning and extinction as a function of anxiety. *J. exp. psychol.*, 1953, **45**, 116–119.
12. SPENCE, K. W., & TAYLOR, JANET A. Anxiety and strength of UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.
13. TAYLOR, JANET A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.
14. TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
15. TAYLOR, JANET A., & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61–64.

Received February 6, 1956.

CONDITIONING OF A RESPONSE CLASS ON A PERSONALITY TEST

ANNE M. NUTHMANN[1]

University of Wisconsin

THE demonstration of verbal learning without awareness of what is being learned has been the subject of much controversy in psychological literature for many years. Recent studies (7, 8, 15) using a word-association technique report conflicting results, but positive findings (7, 8) are of very small absolute magnitude. Studies using other methods (1, 2, 3, 4, 12, 14) that might be expected to involve less interference from the subject's (*S*'s) previous experience have found more striking evidence for learning without awareness. In some (1, 2, 3, 14), verbal stimuli contingent upon *S*'s response have been found to have reinforcement properties, but when mechanical stimuli such as a light or a tone are used, reinforcement sometimes does (4, 12) and sometimes does not (1, 14) occur.

If learning can occur without awareness in areas relevant to personality formation when contingent stimuli are presented in the relatively simple fashion of these studies, the implications for psychotherapeutic methodology are great. A personality variable that has been given much attention in recent literature on psychotherapy is that of "attitudes toward self." Changes from more negative to more positive statements about the self have been shown to occur in successful therapy (9, 10, 11). Also scores on several scales of the MMPI have been shown to change significantly after therapy (5). There has been much less success in correlating therapist behavior with therapeutic outcome (13). The present study is an attempt to demonstrate an increase in responses indicative of acceptance of self on a personality test as a function of the contingent stimuli "good" and a light blink.

METHOD

Selection of items for the personality test. Several hundred items, many from existing personality and interest inventories, the remainder devised by the author, were classified by 80 students of elementary psychology courses at the University of Wisconsin into three categories: acceptance of self, rejection of self, and items having nothing to do with acceptance or rejection of self. Acceptance and rejection of self were defined for the students as follows: "One accepts himself by thinking he is an adequate person, admitting that he has strong points and weaknesses. One rejects himself by thinking he is an inadequate person or by refusing to admit weaknesses which are true of almost everybody." These definitions were written at the top of the list of items which the students were required to classify, but no example items of the three categories were given. An item was included in the experimental test as indicating acceptance or rejection of self if the difference between the proportion of students placing it in one of these categories and the proportion expected by chance was significant at the .001 level. To meet this criterion, it was necessary for 57 out of the 80 students to place it in the same category. The two categories, "acceptance" and "rejection," were then combined since an answer of "false" to a rejection item indicates acceptance. To reduce the probability of the experimental *S*s becoming aware of the conditioning process, buffer items from the third category were added.

The resulting test was administered to a new group of 80 students to determine the relative number of acceptance-of-self answers to each item and to eliminate items that were not answered in the opposite direction from the majority by at least 7 of the 80 students, since these items would be unlikely to be conditioned. The number of "true" and "false" answers indicating acceptance of self were then made equal. The final test consisted of 100 acceptance-of-self items and 40 buffer items.

Selection of subjects. The test was then administered to still another group of 420 students. The *S*s for the experiment were chosen from students scoring in the lowest 15 per cent of the acceptance-of-self distribution. They were matched on the basis of these scores and assigned to one of the three experimental groups, "good," light, or control. There were 15 *S*s in each group. Between one and two months after this pretest the experiment was conducted.

Apparatus. In the small room used for the experiment, the principal piece of apparatus was a 22- × 24-inch fiberboard partition painted flat black and mounted vertically on a table so that it separated *S* and *E*. A 2½- × 4¼-inch opening in the center of the partition served as a display for 3- × 5-inch white index cards containing the items. Immediately above the opening was a hole 1 inch in diameter covered with white paper, which could be illuminated by a 6-v. bulb on *E*'s side of the partition to produce an evenly diffused signal-light to *S*. Lever-type microswitches labeled T (true) and F (false) by means of white paint on the partition were placed 2 inches above the base of the partition and 8 inches in from either side. The switches were wired to 6-v. bulbs on *E*'s side of the partition, and signaled *E* which microswitch had been depressed. The light signaling *S* was controlled by a silent switch that could be set by *E* to go on when the T or F switch was depressed, or not to go on for either. The items, typed on the index cards, were divided into five blocks, each block containing 10 items for which "true" was the accept-

[1] Now at Mental Hygiene Clinic, Great Falls, Montana.

ance-of-self response, 10 items for which "false" was the acceptance-of-self response, and 8 buffer items. The blocks were matched for the number of acceptance-of-self answers to the items within them given by the group of 80 students to whom the test was first administered as a personality test. The order of presentation of the blocks was varied according to a 5 × 5 latin square to control further for possible unequal conditionability of the blocks.

Conditioning. In the experimental session *S* was seated at the table facing the fiberboard partition and was read the following instructions:

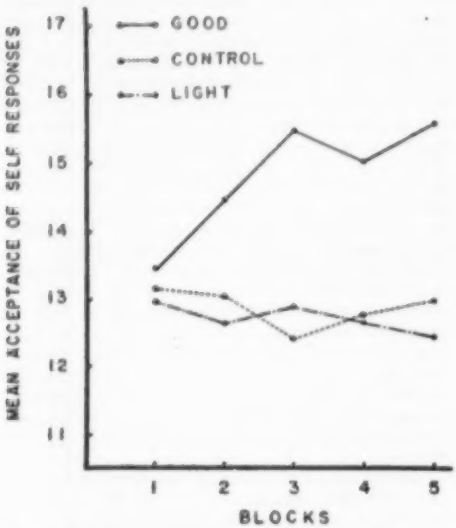


FIG. 1. MEAN ACCEPTANCE-OF-SELF RESPONSES FOR THE THREE GROUPS ACROSS THE FIVE BLOCKS OF ITEMS

TABLE 1

SUMMARY OF ALEXANDER TREND ANALYSIS OF THE NUMBER OF ACCEPTANCE-OF-SELF RESPONSES FOR THE THREE GROUPS

Source of Variation	Error Term	*df*	Mean Square	*F*
a. Between group means	b	2	101.07	3.52*
b. Between individual means	h	42	28.75	8.03***
c. Group deviations from estimation	h	6	.99	—
d. Between group slopes	e	2	20.01	4.03*
e. Between individual slopes	h	42	4.97	1.39
f. Over-all slope	e	1	4.11	—
g. Over-all deviations from estimation	h	3	.47	—
h. Individual deviations from estimation		126	3.58	

* $p < .05$.
*** $p < .001$.

"Several weeks ago in your Psychology 1 class you took a personality test. I am interested in the effect that different ways of presenting personality statements has on the way people reply to them. You will now be given the same test you had before only in somewhat different form. Through this window you will be shown a series of cards on which are typed the personality statements. You are to read each statement carefully and decide whether or not it is true or mostly true with respect to yourself as you feel right at this moment. Then press the appropriate true or false switch in front of you. Hold the switch down for about one second, like this (*E* demonstrates), but do not press a switch unless a card is showing in the window. Try not to take too long to decide on your answers and don't worry about being consistent. It is a very common occurrence for people to change their answers on personality tests when they take them a second time. I am interested in your first impressions right now. I am not keeping a record of your name as I am interested only in group results, so please be as honest as you can in your responses. Are there any questions?"

The *E* then began to present the cards, the rate of presentation being determined by *S*'s response rate. For the two groups to be conditioned, whenever *S* gave an acceptance-of-self response it was followed immediately by the appropriate contingent stimulus. For all groups *E* then removed the card, recorded the acceptance-of-self response, and presented the next card. For the light group *E* set the light-control switch immediately before presenting each card.

After all 140 cards had been presented, *S*s in the light group were asked the following question, "What did you think the purpose of the light was?" The *S*s in the "good" group were asked, "Did you think that anything I said had anything to do with your responses?" On the basis of *S*s' answers to these questions *E* determined awareness of the conditioning process.

RESULTS

The principal findings of the experiment are shown in Fig. 1, in which the mean number of acceptance-of-self responses for the five successive blocks of items is plotted for each of the groups. The distributions within the various subgroups were moderately skewed, and when the data were subjected to Bartlett's test for homogeneity of variance of the *S*s within groups, a chi square of 7.52 was obtained. For 2 *df* this is significant between the .05 and .02 levels, indicating moderate heterogeneity of variance. Lindquist (6) suggests that nonnormality and heterogeneity of variance do not have an appreciable effect on the validity of the *F* ratio unless these deviations are marked. Therefore an Alexander trend analysis was carried out on the data of the three groups after a Bartlett's test of the slope terms failed to show heterogeneity of variance (chi square of 3.29 for 2 *df*). The results of the trend analysis are summarized in Table 1. The

F of 3.52 for Between Group Means indicates that the means of the three groups differed to an extent significantly greater than could be accounted for by individual differences. The *F* of 4.03 for Between Group Slopes was also significant beyond the .05 level. Inspection of Fig. 1 indicates that both of these differences are almost entirely due to the progressive increase in acceptance-of-self responses for the group receiving the contingent stimulus "good." The curves for the light and the control groups are very similar over all five blocks, indicating that the light was not a reinforcing stimulus.

With respect to the performance level of the three groups at the end of Block 1, it is interesting to note that although the groups were matched on the basis of performance on the pretest before the beginning of the experiment it was not possible to match them exactly. The average score for *S*s in the control group on the first test was slightly higher than that of the other two groups. The "good" group overcomes this handicap in the first block of items, suggesting that conditioning takes place from the beginning. However, though it is evident that the "good" group as a whole increased in acceptance-of-self responses over the five blocks, not all of the *S*s in this group did so.

A latin-square analysis of the data was also carried out to determine whether the individual blocks differed significantly from each other. The *F* of 5.98 for blocks was significant at the .001 level, and indicates that certain blocks of items were easier to condition than others, in spite of the fact that the items were originally assigned to the blocks on the basis of equal popularity.

In the "good" group five *S*s, and in the light group four *S*s, became aware of, i.e., could verbalize, the purpose of the experiment. It is fairly certain that those who could not verbalize the correct purpose actually were unaware, because for the most part these *S*s verbalized a contradictory hypothesis, many of them stating that reinforcement occurred whenever their response was consistent with that of previous testing. Since no over-all increase occurred for the light group, only the results of the "good" group were analyzed further. An Alexander trend analysis of the data for the aware and unaware *S*s of the "good" group showed that neither the difference between the group means nor the difference between the group slopes was significant, the *F* ratio for the slope difference being less than 1.00. The aware group, however, did respond at a slightly higher level than the unaware group at all points.

Since the *S*s for this experiment represented one extreme of a distribution, one might expect a regression toward the mean to occur for all the groups on a second testing. There actually was an over-all average increase in total acceptance-of-self responses from the pretest to the experimental test. In the control group this average total increase was 6.74 responses, in the light group 6.53 responses, and in the "good" group 16.86 responses. In the "good" group 7 of the *S*s increased 20 responses or more; 11 of them increased 10 responses or more.

DISCUSSION

The results of the present experiment demonstrate that it is possible to learn without awareness, defined in terms of verbal report, to respond in a more self-accepting way to items on a personality test. Here, as in other experiments on learning without awareness, one may ask exactly what it is that the *S*s learned. In the studies (3, 4, 12) where plural words were conditioned, for example, did the *S*s learn the concept plurals or did they learn to say words ending with the *s* sound? In the present case it is obvious that the *S*s did not learn to respond in a certain way to specific items, because each item occurred only once. They did not learn a "true–false" pattern of responding because the items to be conditioned did not appear in such a systematic order. Also a given order of presentation of the blocks cannot account for the group increase in acceptance-of-self responses since this order of presentation was varied in a latin square design. Another possibility arises as to the effect of differential wording of the items to which "true" and "false" were the acceptance-of-self responses. Very few of the rejection-of-self items were worded negatively, so that the *S*s could not learn to respond "false" to negatively worded items. It is true, however, that most of the times for which "false" was the acceptance-of-self response did contain words that might be classified as being negative in emotional tone, whereas the "true" acceptance-of-self items in general contained words

of positive emotional tone. Whether emotional tone or the broader concept of acceptance of self was actually discriminated remains open to question. In favor of the latter alternative is the consideration that if discrimination of emotional tone did occur, some of the *Ss* should have been able to formulate it verbally, since some of them were able to verbalize the more abstract concept of acceptance of self. For the purpose of this discussion, therefore, it is assumed that discrimination was in terms of self-acceptance.

The verbal stimulus "good" was shown to be capable of bringing about this increase in acceptance-of-self responses. There were, however, large individual differences in the amount of conditioning. Two factors may be suggested to account for these differences. As a result of a differential degree of past learning by the *Ss* of responses involving their self concept, it may take longer for reinforcing stimuli to bring about a measurable change in some *Ss*' behavior than in the behavior of others. That is, *Ss* who have responded, overtly or covertly, "I am worthless" hundreds of times may take longer to change this response than *Ss* who have thus responded only a few times. Also the stimulus "good" may have different reinforcing value for different *Ss* as a result of their previous experiences with it.

The behavior of *Ss* in the group who were aware of the *E*'s intent, it has been noted, did not differ significantly from the behavior of those who were unaware. Four of the five "aware" *Ss* volunteered the additional information that they were quite sure that they had not been influenced by the reinforcement.

The results obtained with the verbal stimulus "good" may have important implications for psychotherapy. It is possible that some of the changes in a patient's verbalized self concept that occur in therapy are the result of simple verbal reinforcement given, perhaps sometimes unknowingly, by the therapist.

The light blink, however, clearly was not a reinforcing stimulus in this situation, not even for the four *Ss* who were aware of the purpose of the light. This finding offers evidence to support a previously suggested hypothesis (14) that a situation of the clinical sort is not conducive to conditioning by means of a light. Perhaps in order to function as a reinforcer of verbal behavior a stimulus must in some way be familiar to the *S* as having a connection with his responses. In the studies that found the light to be reinforcing (4, 12) *Ss* were not responding with meaningful verbal material, but were simply saying words. Presumably the *Ss* had never experienced this type of situation previously, and therefore a blinking light would not be a strange resultant of their behavior. On the other hand, in the studies obtaining negative results with a light (1, 14) and in the present study, the *Ss* certainly had much previous experience with using meaningful sentences but most probably had no experience with being confronted with a light blink as a conveyor of feedback.

SUMMARY

The purpose of the present experiment was to ascertain the reinforcement value of the verbal stimulus "good" and a light blink stimulus on a class of behavior designated as acceptance-of-self responses on a true-false personality test. Each item in this test that was to be conditioned was chosen on the basis of agreement among student judges significantly different from chance at the .001 level that the item measured acceptance of self. The test was then administered to 420 psychology students, and *Ss* for the experiment were chosen from a subpopulation receiving low acceptance-of-self scores.

The *Ss* were assigned to one of two experimental groups or a control group after having been matched on the basis of their performance on the pretest. In the experimental situation the items were presented to the *Ss* individually on index cards. For *Ss* in the experimental groups, each true or false answer indicating acceptance of self was reinforced by *E*'s saying "good" (in one group) or by an immediate light blink (in the other.) In the control group no feedback of any kind was given. The *S*'s awareness of the conditioning procedure was determined from his verbal report at the end of the experimental session.

The results indicate that (*a*) learning without awareness to respond positively to acceptance-of-self statements can occur; (*b*) the behavior of *Ss* who are aware of the purpose of reinforcement does not differ significantly in this situation from that of *Ss* who are not aware; (*c*) the verbal stimulus "good" is capable of bringing about such learning whereas the nonverbal light stimulus is not.

REFERENCES

1. BALL, R. S. Reinforcement conditioning of verbal behavior by verbal and non-verbal stimuli. Unpublished doctor's dissertation, Univer. of Indiana, 1952.
2. COHEN, B. D., KALISH, H. I., THURSTON, J. R., & COHEN, E. Experimental manipulation of verbal behavior. *J. exp. Psychol.*, 1954, **47,** 106–110.
3. GREENSPOON, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *Amer. J. Psychol.*, 1955, **68,** 409–416.
4. GREENSPOON, J. The effect of verbal and non-verbal stimuli on the frequency of members of two verbal response classes. Unpublished doctor's dissertation, Univer. of Indiana, 1950.
5. KAUFMANN, P. Changes in the Minnesota Multiphasic Personality Inventory as a function of psychiatric therapy. *J. consult. Psychol.*, 1950, **14,** 458–464.
6. LINDQUIST, E. P. *Design and analysis of experiments in psychology and education.* Cambridge: Riverside Press, 1953.
7. PHILBRICK, E., & POSTMAN, L. A further analysis of "learning without awareness." *Amer. J. Psychol.*, 1955, **68,** 417–424.
8. POSTMAN, L., & JARRETT, R. F. An experimental analysis of "learning without awareness." *Amer. J. Psychol.*, 1952, **65,** 244–255.
9. RAIMY, V. Self-reference in counseling interviews. *J. consult. Psychol.*, 1948, **12,** 153–163.
10. SEEMAN, J. A study of the process of nondirective therapy. *J. consult. Psychol.*, 1949, **13,** 157–168.
11. SHEERER, E. T. An analysis of the relationship between acceptance of and respect for self and acceptance of and respect for others in ten counseling cases. *J. consult. Psychol.*, 1949, **13,** 169–175.
12. SIDOWSKI, J. B. Influence of awareness of reinforcement on verbal conditioning. *J. exp. Psychol.*, 1954, **48,** 355–360.
13. SNYDER, W. U. Supplementary correlations relative to the relationship between the use of client-centered techniques and the success of psychotherapy. In *Group report of a program of research in psychotherapy.* State College: Pennsylvania State Col., 1953.
14. TAFFEL, C. Conditioning of verbal behavior in an institutionalized population and its relation to "anxiety level." Unpublished doctor's dissertation, Univer. of Indiana, 1952.
15. WEISS, R. L. The influence of "set for speed" on "learning without awareness." *Amer. J. Psychol.*, 1955, **68,** 425–431.

Received February 14, 1956.

THE GENERALITY OF ATTITUDES TOWARD AUTHORITY AND NONAUTHORITY FIGURES[1, 2]

LEROY S. BURWEN
Research Division, Chicago Tribune
AND DONALD T. CAMPBELL
Northwestern University

THE collection of men into large-scale coordinated organizations typically involves a hierarchical order of authority in which the great bulk of persons are placed between a superior and a subordinate echelon. This common situation gives general importance to the description and diagnosis of any generalized attitude toward authority, if indeed such exists. Furthermore this is one predictive setting in which all available psychological theories seem to agree. Writers with as varied orientations as Freud, Piaget, Rogers, Newcomb, Krech and Crutchfield, etc., all agree in positing attitudes towards parental authority originating in the family situation which generalize and manifest themselves as attitudes toward superiors in later social situations. Piaget, for example, writes as follows:

> Day to day observation and psychoanalytic experience show that the first personal schemes are afterward generalized and applied to many people. According as the first inter-individual experiences of the child who is just learning to speak are connected with a father who is understanding or dominating, loving or cruel, etc., the child will tend (even throughout life if these relationships have influenced his whole youth) to assimilate all other individuals to this father scheme (7, p. 207).

There is also a widespread awareness among clinical psychologists that problems of sibling rivalry repeat themselves in peer relationships in adult life. Whether these attitudes or response generalizations are mediated by stimulus similarity in terms of age and physical appearance; by language habits in which the terms father, old man, and boss are used interchangeably; or by learned stimulus equivalence, role similarity, or other nonspatial configurational attributes, the predictions remain the same. The present study attempts to put such predictions to test.

In terms of methodological orientation, this study attempts to measure a small number of attitude dimensions through a large number of independent measurement techniques. The more common approach in which a large number of traits are measured by a single method leaves the investigator unable to distinguish between meaningful psychological relationships and such irrelevant sources of common variance as "apparatus factors" (9) and "response sets" (3, 4). The approach here taken is congruent with the notion of "construct validity" (8). All the measures introduced are accepted as subject to both random and systematic error. None is set up above the others as a "criterion." Validation becomes a process of mutual confirmation among a series of measures all directed toward the same psychological dimension, but maximally independent in terms of apparatus, conditions of administration, and the like.

METHOD

The data were collected during January and February of 1953 from 155 men assigned to 17 bomber crews at Randolph Air Force Base. The men were in crew assembly training in preparation for action in Korea and had been together for approximately two weeks. The 73 officers were typically men in their late 20's and early 30's with World War II experience. The 82 enlisted men were largely in the 19–23 age range in their first tour of military duty. Several crews were assembled at the same time for the group testing. The two hours of testing represented here was only a portion of the research testing to which they were subjected. In the testing there was emphasis upon the fact that the data were for research purposes and would not become a part of any man's record. The practical anonymity and mutual support of the group situation, plus more general attitudes characteristic of the impending transfer to Korea, resulted in a testing situation characterized in some cases by perfunctory compliance and occasional humorous sabotage of the test purpose. In the end, between 10 and 15 per cent of the answer sheets had to be discarded. In addition to

[1] This study was sponsored in part by the United United States Air Force under contract No. AF 18 (600)-170 monitored by the Crew Research Laboratory, Air Force Personnel and Training Research Center, Randolph Air Force Base, Randolph Field, Texas.

[2] This article is in part based upon a dissertation submitted by the senior author to the University of Chicago in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

the group tests, 57 interviews were conducted individually on a basis of compulsory assignment. Particularly for the enlisted personnel, this setting created a guarded, deferential attitude that was difficult to overcome during the one-hour period involved.

Seven different measurement approaches were employed. Through each of these, as appropriate, attitudes were sought toward: own father; symbolic authority (as might be shown, for example, in responses to the picture of older persons in the TAT); immediate boss (the aircraft commander for most crew members); immediate peers (fellow crew members); and symbolic peers. In all, 21 separate variables were scored. Unless otherwise specified, the number of cases upon which the reliability coefficients are based is 155, except for the interview measures, where the N is 57. Officers and enlisted men have been pooled in the analyses that follow, in order to provide maximum stability for the coefficients computed. Significant officer-enlisted differences were found on only 3 of the 21 measures.

Procedures

1. ***Interview.***[3] The interviews focused upon family background and the attitudes toward the current Air Force situation. The subject (*S*) was drawn out on attitudes toward his father, siblings, present superior officers, and peers. Each of these four attitudinal areas was then scored on a five-point scale for favorability or unfavorability. These ratings were made on what the interviewer inferred the covert attitudes of the *S* to be, rather than simply the overt attitudes expressed. While no direct estimate of the reliability of interview ratings is available, a minimum estimate might be the squared correlation coefficient of the interview variable with other test variables (see Table 1).

2. ***Thematic Apperception Test.*** TAT-like pictures were administered by projection upon a screen while the *S*s wrote out their stories. The assignment to respondents was: "You are to tell a story about this picture which will explain what has happened and how things are going to turn out." Eight pictures were used, three from the Murray set (1, 6BM and 7BM) and five specially designed for Air Force use.[4] Cards were selected so as to elicit attitudes toward Symbolic Authority and Symbolic Peers. Each of these attitude variables was scored on a five-point rating scale ranging from highly unfavorable characterization of the person to highly favorable characterization. Older persons and persons of higher military rank were assumed to be symbolic of authority figures while peer-age persons and persons of lower rank were interpreted as symbolic peers or subordinates. The interjudge reliability for these ratings is represented by a correlation coefficient of .83 between one judge and another on a sampling of 92 stories.

Internal-consistency reliability is difficult to estimate on a free-response test where not every person gives a scorable response to every stimulus. Seventy-one *S*s gave some scorable Symbolic Peer descriptions on two of the cards. The correlation between the ratings received on these two cards was .36, which corrected by the Spearman-Brown formula yields a reliability of .53 for a total score based upon the two cards. Scorable responses for Symbolic Peer also were made to other cards. The mean number of scorable Symbolic Peer responses per *S* for the entire 155 *S*s was 1.8. One hundred and eight *S*s gave scorable responses to presumed authority figures on four of the cards. Correlating the sum of the ratings on two of these cards with the sum of the ratings on the other two yields an r of .38 which corrects to .55 for the reliability of the sum of the four card scores. The mean number of scorable Symbolic Authority responses for the total 155 persons was 6.2; thus the reliability based upon four cards is presumably a slight underestimate. The scoring procedure employed was somewhat more direct than many clinical TAT interpretations. In particular, no use was made of the presumed ego representation in the story, nor of more complicated dynamic interpretations (e.g., such as would involve reaction formation). It was therefore thought desirable to rerate the stories on a more clinical basis, using whatever complex cues might be available. It is important to note here, however, that the stories contain much less complexity than those provided by neurotic persons with middle-class backgrounds. These global clinical ratings are designated by TATg while those based upon scoring the descriptions of specific persons summed over the eight pictures are indicated by TATo.

3. ***Description of Self and Others.*** In this test the respondents were asked to evaluate other persons by indicating whether each of 30 different trait terms applied or did not apply. Eight of the 30 terms were to be marked as applying; eight as definitely not applying; and the rest omitted with a resulting intermediate weighting. The descriptions of father, immediate superior (boss) and a fellow worker were scored in terms of the degree to which favorable terms were used in the description. The reliabilities[5] are as follows: father, .24; boss, .34; peer, .55.

4. ***Photo Judging Test.*** In this test the respondents were asked to characterize briefly persons shown in photographs. It was assumed that the photographs of middle-aged and older persons would be most likely to tap attitudes toward those superior in authority, and that reactions to the photographs of high school age persons would be most likely to tap attitudes toward subordinates and peers. For the present scoring purposes the responses to 20 photographs of men over 45 form the basis for the score on attitudes toward Symbolic Authority. The responses to five photographs of high school boys provided the score labeled Sym-

[3] The senior author, with a background of four years of clinical work, conducted the interviews and scored all measures involving clinical or subjective judgment.

[4] The authors are indebted to Dr. William Soskin for two of these pictures; three were designed by the senior author.

[5] Unless otherwise noted, the reliabilities referred to for this and following measures are internal consistency measures of the Kuder-Richardson type, with the computation using item variances and total score variances, following Gulliksen's Formula 10 (5, p. 223). If the assumptions underlying this coefficient are not met, the reliability is systematically underestimated. It is worthy of note that although the reliability of father description is only .24, it correlates .40 with both the interview rating on father and with the father score obtained from the autobiographical inventory.

TABLE 2
MEAN CORRELATIONS BY TOPIC

Topic		Topic				
		F	SA	B	P	SP
Father	F	.35	.12	.03	.06	.08
Symbolic Authority	SA		.15	.08	.10	.06
Boss	B			.09	.13	.03
Peer	P				.22	.07
Symbolic Peer	SP					.01

bolic Peer. The assignment to the *Ss* was to "Write down what you think the person in the picture is like." Space was left for 2 or 3 sentences, although in general a listing of trait terms was given. These responses were rated as to whether the characterization was favorable, intermediate, or unfavorable. Interjudge agreement in making these ratings was .95 for one pair of judges and .86 for another. Internal consistency reliability for Symbolic Authority was .46, for Symbolic Peer, .29.

5. *Autobiographical Inventory.* This inventory contained both short free-response items and check-list items. The full details of the questionnaire are too long for reporting here, and, as with all instruments discussed in this paper, have been reported fully elsewhere (2). Scoring involved both subjective ratings of free-response items and direct scoring of check-list material. Interjudge agreement on a sample of 100 subjectively evaluated items was .86. There were eight items on the *S*'s relationship to his father and these yielded an internal consistency reliability of .56. Attitudes toward boss on this instrument included present boss and previous supervisors and teachers. There were 16 such items which yielded a reliability of .56. There were 11 items describing relationships with peers, including classmates, childhood friends, and fellow workers, giving a reliability of .55.

6. *Attitude Survey.* The tested personnel were also given an attitude survey developed for the study of morale by the Crew Training Research Laboratory, Randolph Air Force Base. Ten of the items of this test, all of which were statements to be endorsed in one of five levels of intensity, seemed to reflect attitude toward authority in the military situation. These items included competence of superior officers and instructors, and general favorableness of attitudes toward the Air Force as an administrative unit. Scoring the responses with five levels of weighting, the symbolic authority score from this instrument had a reliability of .59. Ten items reflected on attitudes toward members of the crew as peers and had a reliability of .73. The peer items were considerably more homogeneous topically than the Symbolic Authority items. These items represented attitudes toward actual peers in the working situation, while the authority items were of such a general nature that we have classified them as attitudes towards Symbolic Authority.

7. *Sociometric Questionnaire.* The "Position Description Test" developed by the Crew Training Research Laboratory required each crew member to pick those fellow crew members whom he would like to be with in five hypothetical situations, including a crew party, shared work responsibility, bailing out behind enemy lines, etc. Each person's attitude toward boss was based upon the relative proportion of his mentions given to the aircraft commander. (Aircraft commanders as respondents were left out of this analysis, reducing the total number of cases to 138.) Peer scores for the officers were based upon the nominations given the other officers; for the enlisted personnel, the score was based upon the responses given to the other enlisted persons. For boss, the reliability was .75. For peer, the values were .84 for officers and .90 for enlisted men.

RESULTS

Table 1 presents the product-moment correlations among these 20 variables. The variables are arranged according to the a priori attitude groupings and the expected pattern of correlations. The a priori expectations were of generalized attitudes among the measures dealing with father, symbolic authority, and boss and also homogeneity of responses among the various measures focused on peers and symbolic peers. A relative independence between these two major sets of attitudes was expected.

The ubiquitousness of "apparatus factors" (9) is indicated by the fact that for 15 of the 20 variables the highest correlations were obtained between measures of different attitudes based upon the same instrument. For only 3 of the 20 variables is the highest correlation with another measure of the "same" attitude through a different instrument. This "apparatus factor" effect is particularly marked for the Interview, TAT, and Sociometric measure. In Table 1, correlations involving the same instrument have been put in parentheses so that they may be discounted in the interpretation. It is obvious that in general the other correlations are considerably lower than expected.

In spite of the disappointingly low general level of correlation, there is a slight general positive trend. Disregarding the intra-instrument coefficients, the values range from $-.28$ to $+.56$. Of the 169 coefficients not involving the same instrument, 143 are positive, differing from a chance distribution at the .001 level, as tested by the binomial expansion.

To examine this matrix for evidence of specific attitudes toward father, boss, peers, etc., Table 2 has been prepared from Table 1 by averaging by z transformation the correlations within each of the cells, eliminating those based on the same instrument, and weighting by number of cases (which reduces the influence of the interview measures). If specific attitudes exist, correlations among

TABLE 3

COMPARISON OF CORRELATION PAIRS TO TEST THE ASSUMPTION OF STIMULUS GENERALIZATION FROM FATHER TO BOSS

Father Measure	Corr. With	Measure of Boss and Peer Attitudes				Score
		Int.	S & O	A.I.	Soc.	
Int.	Boss	.64	.19	−.06	.19	¼
Father	Peer	.65	.27	.04	.04	
S & O	Boss	.08	.23	−.14	.06	¼
Father	Peer	.09	.21	.00	.08	
A. I.	Boss	.15	.09	−.08	−.03	¼
Father	Peer	.24	.03	.15	.05	
						3/12

variables focused upon the same attitude object should be higher than correlations among measures involving dissimilar attitude objects. The data in Tables 1 and 2 indicate that this requirement is clearly met only for attitudes toward father and toward peers. It is most clearly not met for symbolic peers and boss. Even where met, the level of intercorrelation is much lower than one might have expected.

In attempting to interpret these findings it is well to look more closely at the assumptions involved in giving the different measures similar labels. In anticipating some equivalence among the various measures of attitude toward father we assume *response generalization* or response equivalence to a relatively constant, identifiable stimulus. The same assumption is involved for the measures of attitude toward boss and peer, with the exception of the scores taken from the autobiographical inventory, in which the measure is based upon responses to a number of previous bosses and previous peers. Removing the autobiographical inventory measures for boss and peer should thus improve the extent to which they refer to a single attitude object. So doing raises the average correlation among the measures of peer attitude about .05 points and has a similar effect upon the correlations among measures of attitude toward boss. However, even with this alteration, measures of attitude toward boss still fail to meet the simple requirement that the measures correlate higher among themselves than with measures focused on other attitude objects; even with all autobiographical inventory measures removed, measures of attitude toward boss correlate more highly with measures of attitude toward peer than they do with other measures of attitude toward boss.

The other assumption employed in the labeling of the attitude measures is that of *stimulus generalization*. It is on the basis of assumed stimulus generalization that the symbolic peer attitude measures were given similar labels, and that the symbolic authority attitudes were placed in the same category. On the basis of one of the most widespread hypotheses in the psychology of personality we also assumed generalization from father to symbolic authority and to boss (1, 6, 7). In Table 2, it is clear that the three presumed groups with respect to attitude toward authority (father, symbolic authority and boss) correlate no more highly with each other than they do with the two sets of peer measures. Thus while we did find some evidence of specific attitudes toward father, we find no evidence at all for a generalized "attitude toward authority" which would include in its manifestations attitudes both toward father and toward persons in supervisory positions. This finding is so out of keeping with general theoretical anticipations that it is worth examining in greater detail.

The comparisons just made are somewhat unsatisfactory because of the variability or heterogeneity of the correlation coefficients within the blocks averaged. Not all of the attitudes have been measured with all of the instruments, and the instruments differ in the strength of apparatus factors. In addition, stimulus generalization or habit similarity may be expected on a priori grounds, involving social objects sharing the same temporal and spatial situation, and leading in particular to factors centering around the early home, on the one hand, and the present working situation on the other. The correlation between peer and boss attitudes might, for example, be due to the fact that they are both specific manifestations of a more generalized attitude toward the current working situation. To achieve comparisons in which these factors are held constant, we have extracted from Table 1 specific pairs of correlations that are comparable with regard to the instrument and the situational factors involved. Some of these correlation pairs are shown in Table 3. All

TABLE 1

INTERCORRELATIONS OF ATTITUDE MEASURES

	Topic and measure	Topic and Measure																			
		Father			Symbolic Authority					Boss				Peer					Symbolic Peer		
		Int.	S & O	A.I.	Int.	TATg	TATo	Photo	Surv.	Int.	S & O	A.I.	Soc.	Int.	S & O	A.I.	Surv.	Soc.	TATg	TATo	Photo
F	Interview	.40	.09	(.83)*	.14	.08	−.08	.41	(.64)	.19	−.06	.19	(.65)	.27	.04	.07	.04	.35	.30	−.15	
	Description of Self and Others		.40	.20	.17	.18	.09	.14	.08	(.23)	−.14	.06	.09	(.21)	.00	.05	.08	.11	−.03	.09	
	Autobiographical Inventory			.07	.19	.07	−.08	.13	.15	.09	(−.08)	−.03	.24	.03	(.15)	.05	.05	.08	.12	.07	
SA	Interview				.19	.17	−.04	.51	(.81)	.12	−.14	.37	(.77)	.23	.04	.20	.16	.36	.20	−.05	
	TAT (scored globally)					(.72)	.13	.21	.14	.13	.03	.04	−.05	.08	.12	.16	.01	(.49)	(.30)	.07	
	TAT (scored objectively)						.12	.22	.23	.07	.00	.07	.03	.07	.04	.15	.01	(.32)	(.19)	.01	
	Photo Judging Test							.00	−.12	.07	.01	.01	−.28	.14	.04	.11	.02	.14	.00	(.15)	
	Attitude Survey # 6								.56	.14	.03	.15	.26	.30	.09	(.41)	.11	.11	.03	.00	
B	Interview									−.10	−.17	.40	(.76)	.11	.09	.18	.26	.17	.14	−.02	
	Description of Self and Others										.09	.16	−.03	(.45)	.04	.25	.10	.10	−.02	.05	
	Autobiographical Inventory											.02	−.20	.03	(.15)	.11	−.04	−.09	.01	.04	
	Sociometric												.05	.41	.17	.20	(.58)	.06	−.03	.06	
P	Interview													.23	−.04	.05	.08	.23	.08	.20	
	Description of Self and Others														.20	.31	.35	.01	−.05	.17	
	Autobiographical Inventory															.19	.05	.05	−.03	.20	
	Attitude Survey # 6																.36	.05	.04	.15	
	Sociometric																	.05	.01	.15	
SP	TAT (scored globally)																		(.65)	.06	
	TAT (scored objectively)																			−.04	
	Photo Judging Test																				

* Parentheses indicate correlations between attitude measures from the same instrument.

of the pairs of coefficients in this table represent correlations of some one measure of attitude toward father with measures of boss and peer attitudes from the same instrument. The boss and peer attitude measures are comparable, inasmuch as they represent the same environmental factor and are diagnosed by the same instrument. They differ only in their predicted similarity to father. As indicated in Table 3, interview attitude toward father correlates .64 with interview attitude toward boss, and .65 with interview attitude toward peers. Interview attitude toward father correlates .19 with Description of Self and Others attitude toward boss, and .27 with Description of Self and Others attitude toward peer, etc. Each of the pairs is comparable in terms of the extraneous factors of instrument type and situational sources of variance. If there is stimulus equivalence between father and boss, the correlation with boss within every pair should be higher than the correlation with peer. Similar sets are presented for each of the three approaches to measuring attitudes toward father. The column headed "score" indicates the proportion of instances in which the expected relationship was obtained—in only three pairs out of the twelve, over all. Thus, not only do these carefully selected comparisons fail to support the existence of stimulus generalization from father to boss but also they show a trend approaching statistical significance in the opposite direction.

In a manner similar to the analysis in Table 3, a number of other comparisons are possible. A higher correlation may be expected between attitudes toward father and attitudes toward symbolic authority than between attitudes toward father and attitudes toward symbolic peers. The TAT scored globally, the TAT scored objectively, and the Photo Judging Test all yield measures of attitudes toward symbolic authority and symbolic peers. If correlations between these two variables and the three measures of attitudes toward father used in Table 3 are compared, a successful prediction is achieved in only 4 out of 9 cases.

Through analogous procedures the following additional stimulus generalization hypotheses were tested, with the outcomes indicated: symbolic authority and boss should correlate higher than symbolic authority and peer (10/20); boss and symbolic authority should correlate higher than boss and symbolic peer (7/12); peer and symbolic peer should correlate higher than peer and symbolic authority (8/15); symbolic peer and peer should correlate higher than symbolic peer and boss (7/12). Needless to say, none of the outcomes approaches statistical significance.

DISCUSSION

At the inception of this study it was anticipated that persons bring with them into any social situation a repertoire of attitudes or habits which modify their interaction with the persons they meet. The literature of social psychology and personality supports the expectation that prominent within this repertoire of habits or generalized attitudes are attitudes toward authority derived from previous encounters with authority, especially in the early family situation. Basic to this kind of anticipation is an assumption of a symbolic similarity between father and other kinds of authority and a generalization of attitudes from one to the other. With respect to the existence of any such generalized attitude toward authority, independent of a still more general attitude toward people, or a still more specific attitude toward specific father or specific boss, this study has been entirely negative. Furthermore, it has shown no evidence of stimulus equivalence between symbolically represented authority figures and actual authority figures or symbolically represented peer figures and actual peer figures. These findings are so out of keeping with the general assumptions of current social and personality psychology of all schools that they are worth some special consideration.

The simplest interpretive device in situations such as this is to reject the appropriateness of the instruments used. When the instruments are employed, one at a time, this approach is usually convincing, for in the application of any particular instrument there are many reasons why it is not wholly appropriate to the task at hand. In a multi-instrument approach, such as essayed here, it was certainly anticipated that some tests would prove better than others. However, the general level of confirmation of the hypotheses was so low that discriminations among the tests as to their efficacy or appropriateness

cannot be made. If tests are to be rejected, they must be rejected collectively. While the writers cannot claim any special excellence for the instruments employed, they do feel that the selection represents a considerable range of approaches unlikely to share common faults, and that collectively the battery is probably representative of the types of approaches that students of personality or social psychology would be likely to assume to be relevant for the problem studied. Rejecting the instruments is appropriate only if one is prepared to suggest a number of other instruments that would be superior and to specify the kind of application in which one would anticipate confirmation. Also, a rejection of the instruments is in itself a modification of the theory, inasmuch as it restricts or limits the kind of situations in which the theory is expected to be applicable. There may be stronger grounds for questioning the adequacy of the conditions of test administration, although the moderate obtained reliabilities are evidence against this position. But certainly parallel studies under more desirable testing conditions would be advantageous.

If one does not reject the results on these two grounds, then must one reject the theory? It is unlikely that such a recommendation would have any great impact upon current social psychological thinking, so it is probably better to speak in terms of modification of the theory in the direction of restricting its applicability. A sensible interpretation might be that although this kind of symbolic equivalence or stimulus generalization is appropriate to the autistic thought processes in normal dream states and in the waking states of extreme neurotics and psychotics, it does not interfere with the waking perceptions of normal individuals such as constituted the test population for the present study. This seems the most reasonable interpretation to the investigators, even though one would have expected some slight carry-over of autistic thought tendencies into our test situations.

The original predictions of this study seem consistent with the principle of stimulus generalization in conditioning theory. From Gestalt theory one would also expect generalization based upon configurational similarity from one authority situation to another. Within the framework of Freudian theory, our results might be considered less surprising. Freud posits the similarity of opposites, and if reaction-formation mechanisms were present in our subjects, attitudes toward boss and father might conceivably be reversed in these instances. It seems to the writers, however, that in usual psychoanalytic presentations, our original expectations are regularly entertained. Even though our simple measure of favorableness toward father may be eliciting expression of reaction formations rather than basic attitudes, one would generally anticipate that the same reaction formation would be used by the person to conceal from himself and others his underlying hostility toward his boss; or that the same protestations of rebellion which the boy manifested toward his father would be repeated toward the boss, to cover, for example, persistent underlying dependency needs. To explore this possibility, we have searched the correlation matrix for possible evidence of greater agreement from depth measure to depth measure, and from surface measure to surface measure, than between a depth and a surface measure. No such evidence is to be found. Furthermore, the scatter diagrams for each of the coefficients in Table 1 have been inspected without success for curvilinearity or more complicated patterns that concepts such as reaction formation would predict. Only in a complex analysis of inter- and intra-individual variability have evidences of dynamic personality processes been found. These analyses will be reported later and do not reduce the incongruity between our present findings and common assumptions in the area of personality and interpersonal attitudes.

SUMMARY

Presumed attitudes toward father, symbolic authority, boss, peers and symbolic peers were measured by interview, descriptive trait check lists, autobiographical inventory, thematic apperception test, judgment of photographs, an attitude survey and a sociometric questionnaire. Intercorrelations among the 21 measures so obtained are generally low, except where two or more measures come from the same instrument. In such in-

stances the high correlations seem to reflect apparatus factors rather than psychologically meaningful communalities. The total matrix of correlation is generally positive to a significant degree. Of the five attitude topics, only among attitudes toward father and among attitudes toward peer are the correlations of such a nature as to indicate that a specific attitude focus is to some extent being tapped by independent methods. Even here the degree of confirmation is not high. For attitudes toward father, the average correlation among three separate measures is .35; for attitude toward peers, the average correlation among five independent measures is .22. Evidence for a generalized attitude toward authority which encompasses attitudes toward father, symbolic authority, and boss is totally negative, suggesting the need for reconsideration of the applicability of commonly held theory in this area.

REFERENCES

1. ALLPORT, G. W. Attitudes. In C. Murchison (Ed.), *Handbook of social psychology*. Worcester: Clark Univer. Press, 1935. Pp. 798–844.
2. BURWEN, L. S. A study of attitudes towards authority. Unpublished doctor's dissertation, Univer. of Chicago, 1954.
3. CRONBACH, L. J. Response sets and test validity. *Educ. Psychol. Measmt*, 1946, **6**, 475–494.
4. CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt*, 1950, **10**, 3–31.
5. GULLIKSEN, H. *Theory of mental tests*. New York: Wiley, 1950.
6. MURPHY, G., MURPHY, LOIS B., & NEWCOMB, T. M. *Experimental social psychology*. (Rev. Ed.) New York: Harper, 1937.
7. PIAGET, J. *Play, dreams and imitation in childhood*. New York: Norton, 1951.
8. Technical recommendations for psychological tests and diagnostic techniques. *Psychol. Bull.*, 1954, **51**, (*Suppl. No.* 2, Pt 2) 1–38.
9. TRYON, R. C. Individual differences. In F. A. Moss (Ed.), *Comparative psychology*. (2nd Ed.) New York: Prentice-Hall, 1942. Pp. 330–365.

Received February 16, 1956

GUESSING BEHAVIOR AND AUTISM[1]

CHARLES M. SOLLEY, *The Menninger Foundation*
DOUGLAS N. JACKSON, *The Pennsylvania State University*
AND SAMUEL J. MESSICK, *Educational Testing Service*

WHENEVER individuals are required to guess the occurrence of alternative future events without any specific knowledge of what to expect, they probably rely upon previously established expectancies. As *S* gains more and more experience with these "future" events, he can modify his guessing behavior to accord with the "true" probability of occurrence of such events. The present investigators were interested in *Ss*' initial guesses in a probability learning task where motivational factors might influence *Ss*' guessing behavior. The primary hypothesis was that *Ss* with minimal information would overguess the frequency of occurrence of stimulus events when those events had been previously associated with signs of goal attainment or need-satisfaction. A secondary purpose of these studies was to investigate some semantic aspects of Osgood's mediation theory (9).

The primary hypothesis was derived initially in the following way. First let us assume that some stimulus event has been directly associated with goal-reaching behavior or need-satisfaction; i.e., it has become a sign for need-satisfaction and concepts related to it. Further let us assume that stimulus events that signify goal-reaching and need-satisfaction tend to be perceived as larger, brighter, denser, or more numerous than they really are (1, 2, 3, 4, 5, 6, 13). That is, we are dealing with classes of stimuli that are associated with need-satisfaction, and we consider signs of need-satisfaction to be accentuated in perception. Now, if neutral stimuli are associated with these signs, they might also acquire the same perceptual properties as the signs, and be perceived as larger, brighter, denser, or more numerous than they really are. It would seem to follow, then, that if a neutral stimulus is thus perceived as more numerous, for example, *S* may overguess the frequency of its occurrence in probability learning situations.

In the present study certain stimulus events were selected that could be assumed to be related to need-satisfaction or positive value. In order to choose such stimulus events, reference was made to factor analytic studies of meaning reported by Osgood and Suci (10). These studies were based upon correlations among several bipolar scales, the ends of which are defined by adjectives such as *good-bad*, *strong-weak*, etc. Of the three factors extracted—evaluation, potency, and activity—the evaluative factor appeared to reflect meanings usually associated with need-satisfaction or positive value, since such scales as *good-bad*, *pleasant-unpleasant*, etc., had high loadings. For the purposes of the present experiment the three adjectives *near* (loaded .41 on the evaluative factor), *high* (.59), and *bright* (.69), were selected as signs of goodness, pleasantness, and need-satisfaction, and their opposites, *far*, *low*, and *dark*, were selected as signs of unpleasantness, etc.

An experimental manipulation corresponding to each of these pairs of adjectives was also selected—an object moving toward and away from *S* for *near-far*, arrows pointing up and down for *high-low*, and a light going on and off for *bright-dark*. The adjectives and their corresponding manipulations are presented in Table 1 together with the predictions about guessing behavior.

In summary, the relationship between the adjectives and positive value is based upon a factor analysis of meaning by Osgood and Suci (10). The assumption that the experimental manipulations reflect the meanings of the corresponding concepts is partially inferred from previous experiments relating such manipulations to conceptual meaning

[1] This research was completed while the authors were at The Menninger Foundation as USPHS Postdoctoral Clinical Research Fellows of the National Institute of Mental Health. The study was partially supported by a research grant (M-715C) from the National Institute of Mental Health, of the National Institutes of Health, Public Health Service. The authors wish to thank Drs. Gardner Murphy, George Stern, and Riley Gardner for their suggestions in the preparation of the manuscript.

TABLE 1

OUTLINE OF THEORETICAL MODEL LEADING TO PREDICTIONS IN FOUR EXPERIMENTS

Experiment	Experimental Manipulations	Semantic (Adjective) Code	Semantic Associations with Code (Osgood & Suci)	Perceptual Expectancy	Predicted Guessing Behavior
I, II	Object Moving Toward *S* Object Moving Away from *S*	Near Far	Good, Pleasant, Valuable Bad, Unpleasant, Worthless	Many Few	Overguess Underguess
III	Arrow pointing up Arrow pointing down	High Low	Good, Pleasant, Valuable Bad, Unpleasant, Worthless	Many Few	Overguess Underguess
IV	Light Turned On Light Turned Off	Bright Dark	Good, Pleasant, Valuable Bad, Unpleasant, Worthless	Many Few	Overguess Underguess

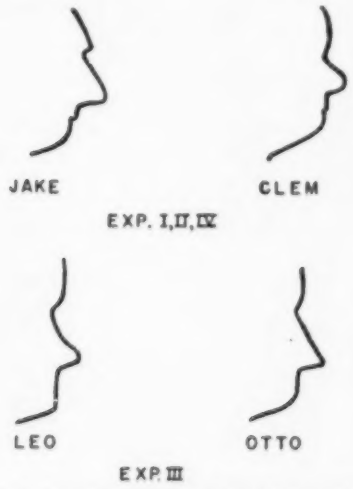


FIG. 1. PROFILES USED IN FOUR EXPERIMENTS

(12). The experimental manipulation is thought to represent a meaning similar to that of the corresponding adjective, e.g., "movement toward" signifies "near," and the manipulation is thus related to need satisfaction and positive value via the concept's semantic relation to similar pleasant and positive meanings.

METHOD

Subjects. In Exp. I, the 17 *S*s were housewives, businessmen, secretaries, and psychologists. Nine *S*s were assigned to a condition in which objects were moved toward *S* when one profile appeared and away from *S* when another profile appeared. The remaining eight *S*s were assigned to a counterbalanced condition in which the movements toward and away corresponded to the opposite profiles. The objects were never given to *S* but were merely moved toward him. In Exp. II the *S*s were 25 female secretarial students, 18 to 21 years of age, from a local business school. Five were assigned to each of four experimental conditions and five to a control group. In Exp. III the *S*s were again 22 female secretarial students, 8 of whom were assigned to one experimental condition and 8 to the other; a control group of 6 *S*s received no experimental treatment. In Exp. IV the *S*s were also 18 female secretarial students, nine of whom were again assigned to each of two counterbalanced conditions. The *S*s in each experiment were randomly assigned to conditions.

Materials. In Exp. I thirty 3 x 5 index cards were used, on each of which appeared a schematic drawing of a right-pointing face. There were two such profiles; one was called "Clem" and the other "Jake" (see Fig. 1). Each profile appeared on 15 of the cards, arranged in a different random order for each *S*. In the experimental procedure one profile was associated with an object moving toward *S* and the other with an object moving away from *S*. The objects used in these movements were quarters, nickles, pennies, salt shakers, paper clips, hairpins, buttons, and burnt matches.

Experiment II involved the same cards used in Exp. I. Only one coin, a half-dollar, and one nonmoney object, a burnt match, were used in the movements toward and away from *S*.

In Exp. III two decks of 30 index cards with similar profile drawings were used. On 15 cards in each deck a profile called "Otto" was drawn and on the other 15 a profile called "Leo" was drawn (See Fig. 1). Again the two profiles were randomly mixed for each *S*. In one set of cards, Deck A, an arrow pointing up was drawn on cards bearing the Leo profile and an arrow pointing down on cards bearing the Otto profile. In the second set, Deck B, the direction of the arrow was reversed for Otto and Leo.

Experiment IV involved the same decks used in Exps. I and II. A small wooden box containing a seven-watt light was used as an association stimulus, the light being turned on and off with a foot pedal. The light was turned on for one profile and off for the other, and the association of a profile with "light on" and "light off" was reversed in a counterbalanced condition.

Procedures. Each experiment consisted of a training series and a guessing series. In the training series *S* was told to watch while *E* turned the cards over one at a time, exposing a profile and saying its name aloud. Then in Exp. I and II *E* moved an object toward *S* during the exposure of one profile and away from *S* during the exposure of the other. In Exp. III *E* merely

turned the cards over, named the profile aloud, and went on to the next card. In Exp. IV *E* turned the light on when one face was exposed and off when the other face appeared.

In the guessing series of each experiment *E* shuffled the deck and then instructed *S* to guess which of the two faces would appear on the next card. When *S* had made his guess, *E* would turn the card over exposing the profile, and *S* would make a guess about the next card. The same deck was used for both the training and guessing phases of each experiment except Exp. III, in which case a deck containing the same profiles but without arrows was used for guessing.

RESULTS

Experiment I. An analysis of variance was done on the number of guesses each *S* made for the profile associated with "toward" movements. This 2 x 2 analysis tested whether the movement of "money" objects had a greater effect than the movement of "nonmoney" objects and whether there were any differences between profiles. The results indicated no significant differences between money and nonmoney, between the effects of the two profiles, or for the interaction of the two variables.

Since there were no significant differences among these experimental variables, results for the different groups of *S*s were pooled to test for differences between guesses about toward- and away-associated profiles. The two means and the *t*-test between them are shown in Table 2. This table indicates that *S*s overguessed the occurrence of toward-associated profiles at the .01 level of confidence. Fourteen of the 17 *S*s tested were in the predicted direction with only three in the opposite direction.

Experiment II. The analysis of this study parallels that of Exp. I. Sixteen *S*s overguessed the profile associated with toward-movements, 3 *S*s underguessed, and there was 1 tie. Again, an analysis of variance on the number of guesses for the toward-associated profiles showed no significant difference between movement of money and nonmoney objects, between profiles, or for the interaction of money-nonmoney and profiles. Again results were pooled over these experimental conditions to test differences between toward- and away-profiles. The means for these two variables and the *t* test between them are shown in Table 2. The toward-associated profile was significantly overguessed at the .05 level of confidence. There was no significant difference between the number of guesses for the two profiles in the control group.

TABLE 2
FREQUENCY OF GUESSES FOR PROFILE ASSOCIATIONS IN RELATION TO CONTRASTING MANIPULATIONS

Experimental Manipulation	Mean	σ_D	t	p
Expt. I (N = 17)				
Toward	16.88	.968	3.88	< .01
Away	13.12			
Expt. II (N = 20)				
Toward	16.00	.788	2.54	< .05
Away	14.00			
Expt. III (N = 16)				
Arrow-up	16.31	.871	3.01	< .01
Arrow-down	13.69			
Expt. IV (N = 18)				
Bright	16.33	1.073	2.48	< .05
Dark	13.67			

Experiment III. Since no significant differences were found in a *t* test between "Leo" and "Otto" profiles, data for both groups were pooled to test the significance of the difference between "arrow-up" and "arrow-down" associations. The *t* test between these two means (Table 2) is significant at the .01 level in the direction of overguessing the "arrow-up" associated profiles. Twelve *S*s overguessed the frequency of occurrence of the profile associated with the arrow up, 2 *S*s underguessed it, and there were 2 ties. There was no significant difference between the number of guesses for the two profiles in the control group.

Experiment IV. Again no significant differences were found between the "Clem" and "Jake" profiles, and the data for the two counterbalanced profile groups were pooled to test the effect of "light-on" and "light-off" associations. The means for these two associations are shown in Table 2 with the corresponding *t* test. Again it was found that *S*s significantly overguessed the occurrence of profiles associated with the light-on condition. Ten *S*s overguessed the profile associated with "light-on," 4 *S*s underguessed it, and there were 4 ties.

DISCUSSION

A discussion of the results and their implications can take several directions. Let us consider the results first in terms of the possible operation of autistic factors and then in terms of their implications for Osgood's mediation theory of meaning (9).

The results indicate that *S*s made more

guesses for profiles associated with movements toward them, arrows pointing up, and lights turning on. These results are interpreted as indicating that the perceptual expectancy for the occurrence of profiles has shifted from equal expectancy, as shown in the control group, to a preponderance for the positively-associated profiles in the experimental groups. If one considers perceptual expectancy to be related to cognitive processes, then this interpretation is similar to the definition of autism given by Levine, Chein, and Murphy (7): "Autism is the movement of cognitive processes in the direction of need-satisfaction."

The *movement* of cognitive processes in the present experimental situation refers to the shift in perceptual expectancies as reflected in the preponderance of responses for positively-associated profiles. This "movement" differs from the type considered in previous autism studies (6, 7, 11), in which attempts were made to elicit *perceptual* shifts in the direction of rewards and away from punishments. The above conception of autism does not completely describe the results of the present experiments, since it is difficult to relate "arrows pointing up" and "lights going on" to need-satisfactions directly. In order to describe the results adequately in terms of an autism theory, a broader conceptualization is required. Some qualification of the phrase "in the direction of need-satisfaction" seems called for. One possible reformulation might define autism as the movement of cognitive processes in ambiguous situations in the direction of meanings or interpretations which have positive value for the organism. Of course, operational specification of the conditions under which cognitive processes "move" and what is meant by "movement" in individual experiments should also be made. In the present studies the movement is identified with a shift in response frequencies in favor of positive associations.

An alternative and not necessarily conflicting interpretation of the present results can also be made in terms of differential emphasis; i.e., movements toward *S*, arrows pointing up, and lights going on might be considered to have a greater emphasizing effect than the opposite manipulations. However, the use of an explanation in terms of emphasis requires some basis for differentiating the relative emphatic value of arrows pointing up v.s arrows pointing down, etc. The hypothesis nevertheless seems amenable to experimental appraisal.

If it is correct to conceive of the association stimuli used in the present experiments as translations of semantic codes into experimental manipulations, then the results also support Osgood's mediation theory and are consistent with his factor analysis of meaning (10). The evidence presented here suggests a behavioral reality for the semantic relationships obtained from Osgood's judgmental studies. Thus, when a scale has a significant loading on the evaluative factor, for example, the verbal and perceptual signs associated with it appear to function evaluatively in behavior.

SUMMARY

As a test of predictions derived from Osgood's (9) theory of meaning and Murphy's (8) theory of autism, four experiments in guessing behavior were conducted. Adjectives highly correlated with goodness and positive value (10) were selected, and translations of these semantic codes into experimental manipulations were specified. Movements of objects toward and away from *S* were chosen to correspond to ***near-far***, arrows pointing up and down to ***high-low***, and a light going on and off to ***bright-dark***. A random series of two profiles was presented to *S*, each profile being associated with one of the experimental manipulations. Later a different random series of the same two profiles was presented, and *S* was asked to guess the order in which they would appear. Profiles associated with operations reflecting the adjectives ***near***, ***high***, and ***bright*** were significantly overguessed while their opposites were underguessed.

REFERENCES

1. ALLPORT, F. H. *Theories of perception and the concept of structure.* New York: Wiley, 1955.
2. ANSBACHER, H. Perception of number as affected by the monetary value of the objects. *Arch. Psychol.*, 1937, No. 215.
3. BRUNER, J. S., & GOODMAN, CECILE G. Value and need as organizing factors in perception. *J. abnorm. soc. Psychol.*, 1947, **42**, 33–44.
4. BRUNER, J. S., & POSTMAN, L. Symbolic value as an organizing factor in perception. *J. soc. Psychol.*, 1948, **27**, 203–208.

5. GILCHRIST, J. C. & NESBERG, L. S. Need and perceptual change in need-related objects. *J. exp. Psychol.*, 1952, **44**, 369–376.
6. JACKSON, D. N. A further examination of the role of autism in a visual figure-ground relationship. *J. Psychol.*, 1954, **38**, 339–357.
7. LEVINE, R., CHEIN, I., & MURPHY, G. The relation of intensity of a need to the amount of perceptual distortion: a preliminary report. *J. Psychol.*, 1942, **13**, 283–293.
8. MURPHY, G. *Personality: A biosocial approach to origins and structures*. New York: Harper, 1947.
9. OSGOOD, C. E. The nature and measurement of meaning. *Psychol. Bull*, 1952, **49**, 197–237.
10. OSGOOD, C. E., & SUCI, G. J. Factor analysis of meaning. *J. exp. Psychol.*, 1955, **50**, 325–338.
11. SCHAFER, R., & MURPHY, G. The role of autism in a visual figure-ground relationship. *J. exp. Psychol.*, 1943, **32**, 335–343.
12. SOLLEY, C. M. Problem-solving difficulty as function of deviation of "meaning" of physical cues from expected "meaning." *J. gen. Psychol.*, in press.
13. SOLLEY, C. M., & LEE, R. Perceived size: Closure versus symbolic value. *Amer. J. Psychol.*, 1955, **68**, 142–144.

Received February 20, 1956

ACCURACY OF INTERPERSONAL PREDICTION AS A FUNCTION OF JUDGE AND OBJECT CHARACTERISTICS[1]

BELA O. BAKER AND JACK BLOCK

Institute of Personality Assessment and Research, University of California, Berkeley

THE processes involved in identifying the reactions and characteristics of other people have been of persistent interest to psychologists for both theoretical and practical reasons. Many recent studies in this area have employed the technique of interpersonal prediction to secure an objective measure of accuracy of understanding. Largely neglected in such studies has been the influence of the person whose behavior is predicted, although it is widely acknowledged that the nature of the social object is an important determiner of accuracy.

The chief aim of the present study was to see whether the level of accuracy achieved in an interpersonal prediction task is predictably affected by systematically selecting the objects of prediction. To this end the objects used were selected to represent three levels on each of two dimensions expected to be relevant to accuracy of prediction.

The first independent variable was *favorableness of self-description*, by which is meant the degree to which an individual portrays himself in a socially desirable light. This variable has been noted previously by Spanner (16) and by Stone and Leavitt (17) as influencing accuracy scores. Spanner found that the most self-critical of four volunteer social objects was less accurately predicted than the other three. Stone and Leavitt employed what they term an "intermediary key" to partial self-favorableness out of the correlation between predicted and actual responses to a personality questionnaire. They found that both general level of accuracy and generality in the sense of correlations between scores achieved for different objects were decreased. Their explanation is that a judge who tends to attribute favorable self concepts to other people will score a number of adventitious hits when predicting the responses of an object who describes himself favorably, regardless of the degree of actual understanding he has of the object. These findings clearly suggest the hypothesis that social objects who are most self-favorable will be most predictable while social objects who are least self-favorable will be least predictable.

The second independent variable studied was *ego control*. Ego control refers to the individual's characteristic means of modulating or coping with his internal needs or drives. Subjects characterized by three kinds of ego-control systems were represented in the present study: Over-controlled (OC), Appropriately-controlled (AC), and Under-controlled (UC). Descriptions of the behavioral properties of these personality systems are given elsewhere (1, 4, 5, 6). Summarizing briefly, OCs bind their tensions excessively and are relatively constricted, categorical, and repressing individuals; UCs tend toward direct and unmodulated expression of their impulses and are relatively expansive, distractable, loosely organized individuals; and ACs selectively bind and gratify their needs in accord with the social reality in which they function.

If it is assumed that the greater the consistency or channelization of an individual's behavior, the more completely he is predictable,

[1] This research was supported in part by the United States Air Force under Contract No. AF 18 (600)-8, monitored by Technical Director, Detachment #7 (Officer Education Research Laboratory), Air Force Personnel and Training Research Center, Maxwell Air Force Base, Alabama. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United State Government.

Personal views or opinions expressed or implied in this publication are not to be construed as necessarily carrying the official sanction of the Department of the Air Force or of the Air Research and Development Command.

The study was carried on at the Institute of Personality Assessment and Research, University of California, and was also supported in part by the grant of the Rockefeller Foundation to that Institute. Since assessment is a group endeavor, it scarcely needs to be indicated that the authors are greatly indebted to the staff of the Institute who conducted the two assessments from which the data are here drawn and who also contributed many helpful suggestions concerning the design and presentation of this research. We are grateful especially to Drs. H. G. Gough, R. B. Holt, and F. Vanasek who made detailed criticism of an early draft of this manuscript. Many of their suggestions have been incorporated in this final version.

then the rank order of accuracy scores for these three groups of objects can be predicted. OCs, who are conceptualized as developing relatively set ways of behaving, should be most predictable; UCs, who are characterized as manifesting frequent inconsistencies, should be least predictable; while ACs should be predictable at an intermediate level.

A further hypothesis was made concerning the accuracy of judges classified in terms of ego control. Our expectation was that OC and UC judges would be less accurate than AC judges. Our assumption in making this prediction was that two components are required for accurate social prediction, sensitivity of social perception, and the ability to integrate perceptions into a formulated prediction scheme. OCs are considered to be relatively excluding of the potential social field while UCs, because of their inability to establish structure or a frame of reference for themselves (4), are relatively unable to construct a coherent basis for prediction. Since ACs possess both required characteristics, they should be superior to both OCs and UCs.

METHOD

Stimulus material available to judges. In the Spring of 1953, 100 Air Force captains were studied during a three-day assessment program at the Institute of Personality Assessment and Research. A standardized stress interview, conducted by a grim and unresponsive examiner, was the initial procedure experienced by each officer. The interview was conducted under floodlights before a group of observers with a sound movie camera operating[2] in plain view of the subject. The schedule of questions started with, "How does it feel to be in the movies?" and moved on to such questions as, "What sort of a person are you?," "What do you consider to be your main personality assets?," and "What do you consider to be your main personality handicaps?" The questions elicited rather rich samples of the expressive and coping behavior of the subject.

Nine officers were selected from the pool of 100 potential social objects on the basis of the considerations indicated below and the films of the interviews served as the standard stimulus material in the study. The films were approximately two and one-half minutes in duration and were of good quality both visually and with respect to sound reproduction.

Nature of the required prediction. In the course of the assessment, each officer Q-sorted a set of 54 adjectives along a 9-point scale so as to describe himself. The task set for the judges was to predict these self-descriptions. To make the judges' task somewhat lighter, several categories were combined, reducing the sorts to a 5-point scale with the fixed frequency distribution: 2, 10, 30, 10, 2. The accuracy score employed was the correlation coefficient between the judge's predicted Q sort for an object and that object's self-sort.

The self Q sort was chosen for prediction because the content of the movies appeared to be highly relevant to the task of evaluating the object's self-percepts and because the prediction of some form of self-description is a widely used experimental task. An important methodological advantage is that the response set variations in elevation and dispersion which Cronbach (9) has noted as confounding the discrepancy score are eliminated by the use of a forced distribution.

The Objects

Measurement of object self-favorableness. To assess self-favorableness, six psychologists were requested to sort the 54 adjectives to predict the self-description of an individual characterizing himself in the most favorable light. The average intercorrelation among the six maximally self-favorable sorts was .73, indicating that a composite of these six Q sorts has an estimated reliability (Spearman-Brown correction) of .94.

This maximally self-favorable Q sort was employed as a reference key with which the self-sorts of the 100 officers were correlated. A high correlation implied an officer was describing himself favorably; a low correlation that he was ascribing socially undesirable characteristics to himself. Selection of social objects in terms of self-favorableness was based upon this set of correlations treated as scores.

Measurement of ego control among the objects. At the end of the three-day assessment program, eight staff psychologists codified their personality evaluations of each of the officers by means of a 76-item Q sort. The eight separate Q sorts for each subject were combined into one composite Q sort which thus represented the staff consensus as to the nature of the subject's personality. The sample of officers was divided into equivalent groups of 50 and, within each group, the consensus Q descriptions were intercorrelated. These correlations between persons were then cluster-analyzed (18).

A number of personality "types" was found, five of which were observed in both 50 × 50 analyses. The cross validated types were combined. When evaluated in terms of test and personality scale measures, the cluster-types isolated by the inverse analyses proved to be reliably different among themselves, and in a fashion compatible with the defining characteristics of their

TABLE 1

CORRELATIONS BETWEEN SELF Q SORTS AND THE FAVORABLENESS KEY FOR THE NINE SOCIAL OBJECTS

Self-favorableness	Ego Control		
	Over-Controlling Objects	Appropriate Controlling Objects	Under-Controlling Objects
High Self-favorableness	.69	.61	.64
Middle Self-favorableness	.42	.43	.36
Low Self-favorableness	.08	.08	.03

[2] Dr. Victor B. Cline served as cameraman during these interviews.

modal Q sorts. A fuller explanation of the procedure and results of these analyses is presented elsewhere (2).

On the basis of the Q-sort descriptions and other measures characterizing them, three of the groups were designated as fitting the categories of over-, under-, and appropriate-control (*Ns* of 26, 26, and 21, respectively). These groups constituted the populations from which the social objects were drawn.

Specification of the social objects. Representatives of the three ego-control groups were selected at each of three levels of self-favorableness, making a total of nine social objects. Table 1 shows the degree of matching for favorableness within the three ego-control groups. The ages of the social objects varied from 29 to 44, with a mean of 34.4. Their socioeconomic origins are best estimated as upper-lower or lower-middle class.

The Judges

The sample of judges. During the winter of 1954–55, 70 applicants to medical school underwent an 18-hour assessment at the Institute. These assessees were preponderantly undergraduate, although a number were graduate students or in the nonacademic community. They ranged in age from 19 to 33, with a mean at 22. Thus, the judges were substantially younger than the social objects. Socioeconomically, the group was essentially of lower-middle to middle-middle class origin, somewhat higher than the social objects.

Procedure with respect to judges. The procedure was administered to groups of five judges at a time in two 40-minute sessions. Before viewing any films, the subjects Q-sorted the 54 adjectives so as to describe themselves. They were then instructed and four films were shown, with suitable intervals for the judges to indicate their predictions. After an interval of two hours, the remaining five films were shown. The films were presented in a standard order such that no social object followed immediately after another object from the same ego-control or self-favorableness group.

Measurement of ego control among judges. At the end of the assessment, seven staff psychologists codified their personality evaluations of each subject-judge by means of a 115-item Q set. The seven separate Q sorts describing each subject-judge were combined into one composite Q sort representing the staff consensus as to the subject's personality.

Two of the Q items that were evaluated by the assessors for each subject were, "Tends toward overcontrol of his needs and impulses; binds his tensions excessively; delays gratification unnecessarily," and "His various needs tend toward relatively direct and uncontrolled expression; unable to delay gratification; acts out." The relative weighting of these two items provided the basis for categorizing judges into one of the three ego-control groupings. For each subject-judge, his Q score on the under-control item was subtracted from his Q score on the over-control item. A distribution of difference scores was thus obtained where high positive scores implied over-control and negative scores under-control.

Four judges were eliminated from the sample for various reasons: one because he grew up in a foreign country, one because he was extremely deviant from the rest of the sample on a variety of assessment measures, and two because they appeared to belong more clearly to an ego-control category not here used—that of bimodal control. The remaining 66 judges were divided into three groups on the basis of the distribution of difference scores. The 22 individuals with the highest positive difference scores were designated as the group of over-controlling judges; the 22 individuals with the highest negative difference scores constituted the group of under-controlling judges; and the 22 remaining subjects in the midrange were classified as appropriately-controlling judges.

Although the judges were categorized on the basis of only two Q items, other items which covaried with these two could be examined to give a more comprehensive picture of the differential characterizations of the groups. It was not possible to make a direct comparison of the Q-sort descriptions of the judges and objects, since different Q decks were used. Inspection of the items differentiating among the three judging groups yields a subjective impression that the dimension involved is quite comparable to that distinguishing among the object groups.[3] The OCs are in both cases described as relatively constricted, passive, conforming individuals; the UCs as more ascendant, expressive, and rebellious; and the ACs as more moderate individuals who share the desirable but not the undesirable traits of the extreme groups.

RESULTS

Accuracy was scored by correlating predicted Q sorts with the appropriate object self-sorts. The accuracy scores were then treated by an analysis of variance design[4] adapted from

[3] Tables of the items significantly differentiating among UCs, ACs, and OCs in the samples of objects and judges have been deposited with the American Documentation Institute. Order Document No. 5035 from ADI Auxiliary Publications Project, Photoduplication Service, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to chief, Photoduplication Service, Library of Congress.

[4] It should ge noted that there is a problem involved in applying analysis of variance to these data. For practical reasons it was necessary to present the stimulus materials to the subjects in a fixed, rather than random, order. Consequently, the possibility that differences between object groups are confounded by order effects must be considered. The order used placed one member of each object classification in each third of the presentation series. Thus the ordinal positions of objects classified by ego-control were: OCs—1, 6, 8; ACs—3, 4, 7; UCs—2, 5, 9. For objects classified by self-favorableness the ordinal positions were: High—1, 4, 9; Middle—3, 5, 6; Low—2, 7, 8. Our trust was that this approximate balancing of serial position within groups would largely exclude order effects from the main comparisons and the simple interactions.

Since the correlations which comprised the raw accuracy scores appeared by inspection to form a reasonably symmetrical distribution, they were coded for this analysis by simply dropping the decimal point and adding a constant of 30 to remove negative numbers.

TABLE 2
ANALYSIS OF VARIANCE OF PREDICTIVE ACCURACY SCORES

Source	*df*	Sum of Squares	Error Term	*F*
Between personality-style of judges (JP)	2	2,596.46	b	3.05*
Error b	63	26,843.58		
Between subjects	65	29,440.04		
Between personality-style of objects (OP)	2	4,909.12	w_2	6.07†
Between self-favorableness of objects (OF)	2	36,568.50	w_3	94.48†
Interactions:				
OP × OF (or OID-Residual individual differences among objects)	4	151.08	w_4	—
OP × JP	4	253.94	w_1	—
OF × JP	4	958.04	w_3	1.24
JP × OID	8	687.53	w_4	1.68
Error w	504	88,205.56		
Error w_1	126	50,915.38		
Error w_2	126	24,383.25		
Error w_3	252	12,906.93		
Within subjects	528	131,733.77		
Total	593	161,173.81		

* For 2 and 60 degrees of freedom, $F_{.10} = 2.39$, $F_{.05} = 3.15$.
† For 2 and 120 degrees of freedom $F_{.001} = 5.54$, $F_{.001} = 7.31$.

one described by Lindquist (13, pp. 292–297). Table 2 summarizes the results. Two of the main effects were highly significant (p less than .005), while the other was of borderline significance (p approximately .06). None of the interactions was significant.

The social objects expressing the most favorable self-descriptions were most readily predicted, the average accuracy score achieved for them being .42. The objects at the middle and lowest levels of self-favorableness did not differ reliably in predictability. Indeed, the low favorableness objects were somewhat more predictable than those in the middle range, average accuracy scores for the groups being, respectively, .24 and .27. Our hypothesis concerning the effect of self-favorableness on predictability is thus partially supported and partially infirmed.

The over-controlling social objects were most accurately judged, the average accuracy score achieved for them being .35. The appropriate- and under-controlling objects, for whom average accuracy scores were .28 and .29 respectively, did not differ in predictability. Our hypothesis concerning this effect is thus also partially supported and partially infirmed.

The AC judges performed at the highest level of accuracy, achieving an average score of .33, and the OC judges were least effective, achieving an average score of .28. The intermediate UC judges, with average scores of .31, were not significantly different from the judging groups at either extreme. Our hypothesis concerning the judges' relative accuracy thus received some support.[5]

DISCUSSION

Accuracy of prediction as a function of the judge. Many current theories of personality suggest that the ability to understand others is an important feature of good adjustment. From this point of view, one would expect that AC judges, who are conceptualized as responding more adaptively to reality, would be more accurate than either OCs or UCs in their accuracy of prediction. And in fact, this was the case, although the magnitude of the differences is small. Since this finding is in general agreement with previous results (e.g., 8, 10, 12), it will not be discussed further except to point out that the result does not necessarily imply AC judges, or better adjusted judges, are more "understanding" of others. At this stage of knowledge, the result simply indicates that these judges are more accurate predictors of others, regardless of how that accuracy is achieved.

The question can be raised at this point as to whether the various groups of judges differed in the manner in which they made predictions. In exploring this question a number of measures were defined, two of which are mentioned here because of their particular relevance. The first measure, labeled "ascription of self-favorableness," is an index of the degree to which favorable self-opinions are attributed to others, a measure which was defined as the correlation between the composite of an individual's predictions and the favorableness key

[5] It should be noted here that the observed level of significance for this hypothesis is an underestimate. The over-all F-test, in the situation where a prediction is supported, provided too conservative an estimate of the true level of significance. An allowance for the ordering of the three groups as predicted is difficult to decide upon. A halving of the obtained p level (.06) is still somewhat conservative but is closer to the correct significance value.

mentioned earlier. An analysis of variance of these scores indicated that there were no significant differences among the groups of judges.

The second measure, tentatively called "self-generalization," was defined as the correlation between the composite of a judge's predictions and his own self-description. An analysis of variance of these scores again indicated that the differences among the groups of judges were not significant. The variations in accuracy described above, therefore, cannot be explained either in terms of variations in the tendency to ascribe favorable self-opinions to others or in terms of variations in the tendency to ascribe one's own self-perception to others.

Accuracy of prediction as a function of the object. The degree of favorableness with which the social objects described themselves was found to be strongly related to the accuracy with which the judges were able to guess their self-descriptions. However, our hypothesis concerning this effect was not fully supported, for, although the highly self-favorable objects were best predicted, the objects with middle and low degrees of self-favorableness did not differ in their predictability and in fact were ordered differently than had been anticipated.

One partial explanation of this result is to be found in the fact that all the judges tended to predict relatively favorable self-descriptions, the average correlation between the composites of individual judge's predictions and the favorableness key being .39, with a standard deviation of .15. Consequently, the degree to which an object's self-description correlated with the favorableness key helped determine the accuracy with which it was predicted. Such a finding—which is not new with us (cf. 16, 17)—is, upon reflection, to be expected from the context of prediction experiments. There is a good deal of social pressure toward the suppression of unfavorable views of one's fellow man. We may cite here the homily, "If you can't say something good about a person, don't say anything." Some of our subject-judges may have been more than ordinarily aware of this pressure because of the fact that they were undergoing an assessment and may have felt that it would reflect ill upon them if they ascribed unfavorable characteristics to the objects. Others may have reasoned (and not too inaccurately) that the objects themselves, since they were also assessees, would have been concerned about making good impressions at the time they sorted the adjectives. In general, the fact that the experimental situation was ambiguous and even a little threatening from the subjects' point of view is likely to have promoted cautious—i.e., favorable—responses.

Beyond the common tendency toward caution evoked by the role of being a judge, there probably are reliable individual differences in the extent to which judges adopt the approach of ascribing favorableness. These individual differences among judges may well be in themselves fruitful objects of study. But the degree to which they should be allowed to enter into measures of judging ability must be determined by the interests of the investigator. If the research focus is on accuracy of prediction per se, then all mediators of accuracy must be accepted as relevant. If, however, the orientation is not toward predictive accuracy per se, but rather toward a notion of "interpersonal acuity," "social sensitivity," or "understanding," then the judges' tendencies to make favorable predictions may be interpreted as confounding, and this source of variance should in some way then be controlled or partialled out.

De-emphasizing the effects of object self-favorableness may be accomplished in a number of ways. Some form of partial correlation may be employed to eliminate the ascription-of-favorableness effect, as in the Stone and Leavitt (17) study, leaving a reduced score but one free of this particular component. But as perhaps the better technique, judges may be prevented from predicting favorableness by the stratagem of employing a forced choice procedure where the alternatives are matched for social desirability, as in, for example, the Edwards Personal Preference Schedule (11).

The influence of the objects' ego-control style on predictability is not to be so neatly accounted for as is the influence of self-favorableness. Because our hypothesis concerning this effect was only partially supported, the result probably does not warrant extensive discussion. While the greater predictability of OC objects supports the conceptualization of over-control previously advanced, the meaning of the equivalence, even slight reversal, of the predictability of AC and UC objects remains obscure.

Although the finding itself does not clarify the process of understanding, it nevertheless

demonstrates that predictive accuracy scores can be manipulated by varying an inferred central dimension. This demonstration takes on more force when one considers the really startling (at least to us) finding that the residual individual differences among objects—i.e., all individual differences among objects over and above those accounted for by the main effects—failed to elicit any significant variance in accuracy of prediction. Thus by systematically varying social objects in terms of self-favorableness—a dimension defined by explicit features of the objects' self-descriptive responses—and ego-control—a postulated structural characteristic of personality—all the variance in object predictability in this experimental situation has been exhausted.

While this finding impresses us as noteworthy, we would hesitate to make too much of it until it is repeated. Self-favorableness is an obvious and easily controlled characteristic of verbal descriptions, the relevance of which is greatly amplified within the context of social prediction experiments. Ego control is a personality dimension underlying a broad variety of behaviors. That the two together should account for the bulk of the variance in predictability is not implausible. Even if this is the case, however, it still does not exclude the possibility that other dimensions overlapping with the ones used here may bear a significant relationship to predictability.

Favorableness, as indicated earlier, is a conveniently manipulated characteristic of verbal descriptions, whether of self or of others. Such is not the case with ego control as a determinant of predictions. It might be possible to derive a key from the present data to reflect those differences among objects attributable to their different ego-control patterns. This key could then be used to partial out of the accuracy scores those differences among them attributable to the objects' ego-control styles. But clearly, such an after-the-fact key would have little general application, since it could not be defined independently of the objects. Experimental control over the effects of this variable cannot be attained, therefore, by the expedient of using partial correlations.

This fact, of course, poses no problems for studies of predictability and merely complicates the study of the mediation of accuracy by the judges' response tendencies. For studies of judging ability, however, the indication seems to be that careful attention to the problem of sampling is essential if results are to be generalized beyond the specific experimental situation. In this respect, Brunswik's thesis (7) applies—the same considerations that govern the sampling of judges should apply to the sampling of objects.

Dimensional extremeness as a possible determinant of predictability. In the present experiment, the social objects in the midrange of the independent variables studied, i.e., the middle level of self-favorableness and the AC level of ego control, were least accurately predicted. In each case the difference in predictability between these middle groups and the extreme group which fell next in order of predictability was slight and nonsignificant. But the ranking itself makes implausible the assumption that there is an ordinal relation between predictability and either of the experimental dimensions. We would expect that the inclusion of more extreme representatives of low self-favorableness and under-control among the social objects would make the curvilinear trend more pronounced.

These findings, in conjunction with our knowledge of the process of concept formation and the development of frames of reference (14, 15), suggest that judges seek clear and unambiguous categories ("stereotypes") as aids in prediction. Socially valid personality categories, informal though they may be, provide an economical and integrative framework from which specific predictions may be made. "Likes himself"—"doesn't like himself" or "rigid"—"impulsive" are readily available organizing principles with a good deal of social utility.

The social validity of these categorizing principles, however, stems from the behavioral consistency of individuals at extreme positions on the continua used. In predicting the explicit self-descriptions of self-ambivalent objects or in assessing the conditional behavior of appropriate-controllers, simple notions of consistency no longer apply. Consequently, accuracy of prediction suffers.

SUMMARY

A sample of nine social objects was selected so as to be systematically varied on two dimensions: ego control and self-favorableness. Sixty-

six subjects viewed sound movies of the behavior of each object in a stress interview and attempted to predict his self Q sort of 54 adjectives. The judges were also classified on the dimension of ego control. An analysis of variance of the accuracy scores indicated that significant variance was associated with ego-control style of the judges, and with both ego-control style and self-favorableness of the social objects. Residual individual differences among the objects were *not* significant. The results are discussed in terms of their implications for the methodology of social perception studies, with emphasis on the importance of a carefully considered selection of standard social objects.

REFERENCES

1. BLOCK, J. An experimental investigation of the construct of ego-control. Unpublished doctor's dissertation, Stanford Univer., 1950.
2. BLOCK, J. A differentiated approach to the officer selection problem. *Officer Educ. Res. Lab., USAF Pers. Train. Res. Cent. Res. Rep.*, March 1955, Project No. 7730, Contract No. AF18(600)-8 (Rep. No. 8). Pp. 1–34 (dittoed).
3. BLOCK, J. & BLOCK, JEANNE. An investigation of the relationship between intolerance of ambiguity and ethnocentrism. *J. Pers.*, 1951, **19**, 303–311.
4. BLOCK, J., & THOMAS, H. Is satisfaction with self a measure of adjustment? *J. abnorm. soc. Psychol.*, 1955, **51**, 254–259.
5. BLOCK, JEANNE. An experimental study of a topological representation of ego-structure. Unpublished doctor's dissertation, Stanford Univer., 1951.
6. BLOCK, JEANNE, & MARTIN, B. Predicting the behavior of children under frustration. *J. abnorm. soc. Psychol.*, 1955, **51**, 281–285.
7. BRUNSWIK, E. *Systematic and representative design of psychological experiments*. Berkeley: Univer. of California Press, 1947.
8. CLINE, V. B. Ability to judge personality assessed with a stress interview and sound-film technique *J. abnorm. soc. Psychol.*, 1955, **50**, 183–187.
9. CRONBACH, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychol. Bull.*, 1955, **52**, 177–193.
10. DYMOND, ROSALIND F. Personality and empathy. *J. consult. Psychol.*, 1950, **14**, 343–350.
11. EDWARDS, A. L. *The Edwards Personal Preference Schedule*. New York: Psychological Corp., 1954.
12. GAGE, N. L. Accuracy of social perception and effectiveness in interpersonal relationships. *J. Pers.*, 1953, **22**, 128–141.
13. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
14. NEWCOMB, T. M. *Social psychology*. New York: Dryden, 1950.
15. SHERIF, M. *The psychology of social norms*. New York: Harper, 1936.
16. SPANNER, M. Similarity, identification, and distortion as factors in the prediction of personality characteristics. Unpublished doctor's dissertation, Univer. of California, Berkeley, 1954.
17. STONE, G. C., & LEAVITT, G. S. Generality of accuracy in perceiving standard persons. Paper read at Midwest. Psychol. Ass., Columbus, Ohio, May, 1954.

Received February 20, 1956

DIFFERENCES BETWEEN SCHIZOPHRENIC AND BRAIN-DAMAGED GROUPS IN CONCEPTUAL ASPECTS OF OBJECT SORTING

LAURENCE S. McGAUGHRAN[1]
University of Houston
AND LOUIS J. MORAN[2]
University of Texas

SCHIZOPHRENIC thinking has been characterized by a number of investigators (e.g., 1, 2, 7, 8, 11, 14, 18, 20, 21, 22, 23) as an impairment or a deterioration in abstract ability, resulting in a "very great concreteness" (7, p. 23). Many of these same writers (e.g., 1, 2, 7, 11, 14, 23) have emphasized that the extreme concreteness of schizophrenic thinking closely resembles that of the brain-injured.

It is the contention of the investigators that studies of conceptualization have been unduly influenced by Goldstein's dichotomous formulation of "abstract-concrete." The two "conceptual levels," abstract and concrete, were originally employed by Goldstein and Gelb to distinguish the conceptual behavior of brain-damaged patients from that of normal adults. When the term, "concrete attitude," was applied subsequently by Goldstein and other investigators to characterize also the conceptual behavior of a number of groups other than normal adults, the inevitable result was to crowd together a heterogeneous group of referents for the term, without apparent univocality.

In a previous report (16), the investigators suggested that the varied uses of the concepts, abstract and concrete, represent a fusion of at least four isolable conceptual variables: (*a*) amount of social agreement, or communality; (*b*) order (or degree) of conceptual classification; (*c*) relative essentiality, or saliency; and (*d*) the presence or absence of generic terms. In that study we attempted to demonstrate that of these four meanings the one that most clearly distinguished schizophrenic and normal conceptualization was the relative amount of social agreement, or communality, for the concepts employed; on the other hand, hypotheses of group differences which were based upon the Goldstein dichotomous abstract-concrete formulation were not supported in a single instance.

In the present study, the hypothesis to be tested is that the second variable—order of conceptual classification—shows the greatest difference between schizophrenic and brain-damaged *Ss*.

METHOD

Subjects

The major samples consisted of a group of 37 male patients diagnosed as chronic paranoid schizophrenic and a group of 34 male patients who had been diagnosed as suffering from types of disorder associated with the more general diagnostic term, brain-damaged. The specific medical diagnoses included encephalopathy or chronic brain syndromes associated with encephalitis, cerebral thrombosis or arteriosclerosis, central nervous system syphilis, epilepsy, post-traumatic encephalopathy, Pick's Disease, etc.[3] A group of 19 additional brain-damaged male patients whose official diagnoses included references to psychotic reactions were used in this study only to check reliability of scoring in the object-sorting procedure.

Means and standard deviations for the control variables of age, education, and estimated intelligence level are shown in Table 1. The distribution of age and education shows close agreement; neither of the mean differences for these variables approached significance. On the other hand, the significance of group mean differences in intelligence exceeds the .001 level. Since intelligence could not be controlled experimentally, and since the results of a previous study (16) had indicated a high degree of association between intelligence and conceptual behavior, analysis of covariance was employed to provide a statistical control over the effects of intelligence upon sorting behavior.

The Sorting Task and Rationale for Scoring Conceptual Areas

The sorting task used was the Rapaport modification of the Goldstein-Gelb-Weigl Object Sorting Test. The task requires 19 different sortings of a group of 33 objects familiar to everyday experience. Administra-

[1] Employed part-time at the Houston Veterans Administration Hospital, where this study was conducted.

[2] Study conducted while employed at the Houston Veterans Administration Hospital.

[3] We wish to express our deep appreciation to Dr. Ruth M. Hubbard and other staff members of the Waco Veterans Administration Hospital for their kind cooperation in making these records available.

TABLE 1
MEANS AND STANDARD DEVIATIONS OF DISTRIBUTIONS FOR CONTROL VARIABLES OF AGE, EDUCATIONAL LEVEL, AND ESTIMATED INTELLIGENCE

Experimental Group	*N*	Age		Education		Intelligence*	
		Mean	*SD*	Mean	*SD*	Mean	*SD*
Brain damaged	34	40.1	10.5	9.3	3.2	95.8	14.7
Schizophrenic	37	40.3	9.6	10.7	2.6	108.9	15.1

* Based upon verbal IQ scores on the Wechsler-Bellevue Scale; for the schizophrenic sample, these scores were estimated from results obtained with the vocabulary subtest.

tion was in accordance with the procedure described by Rapaport (18).

Each sorting was scored separately as occurring in one of four "conceptual areas." The basis for scoring sortings into the four conceptual areas has been described in some detail in previous reports (15, 16). In summary, the rationale is as follows:[4] Two variables, order of conceptual classification and extent of social agreement, are isolated in the operations of this experiment and are termed the "open-closed" and the "public-private" variables, respectively. In previous studies these variables have been found to be unrelated to each other.

Let us consider, first, the open-closed variable. The more attributes that are "bound-together" within a conceptual classification, the more "closed" the sorting, or concept, is considered to be. Thus, a concept based upon a single attribute common to all objects within the conceptual grouping would be maximally "open" (e.g., a group of various objects conceptualized as belonging to the same class because all are composed of metal). On the other hand, a concept binding together all perceptible attributes (all of the "degrees of freedom") of the objects within a conceptual grouping would be maximally "closed" (e.g., a concept based upon the absolute likeness of two sugar cubes). It can be seen that the more open a concept is, the more abstract, or generic, or the higher the order of the principle of classification that is involved.

As for the second variable, "public-privateness," a relatively high degree of publicness of a concept implies that the principle underlying the grouping is widely shared by and freely communicated to others within a social group. For example, the boundaries for a concept of "roundness" are relatively clearly denotable; the conceptual boundaries for a concept of "goodness" would be considerably less so; thus, the exact conceptual limits intended by the latter term are not easily communicated, nor are they easily anticipated by others.

If these two variables, with their respective poles of closed-open and public-private, are visualized as intersecting in space, the resulting quadrants may be conceived to represent the following four "conceptual areas": closed-public, open-public, open-private, and closed-private. These four conceptual areas are the basic scoring categories. Examples of subcategories and typical conceptual groupings within each of the four conceptual areas are as follows:

1. Closed-public area
   a. Absolute identity—(2 corks; basis: "exactly alike").
   b. Species identity—(real and toy forks; basis: "both forks").
2. Open-public area
   a. Unitary abstraction—(all metallic objects; basis: "all made of metal").
   b. Superordination—(several toys; basis: "all these are toys").
3. Open-private area
   a. Loose-representational—(bicycle bell, cork; basis: "you shouldn't drive when drinking").
   b. Over-abstraction—(red ball, red paper circle, red matchbook, red sink stopper; basis: "manufactured articles").
4. Closed-private area
   a. Species-object mediation—(bicycle bell, lock; basis: "put bell on bicycle and lock it").
   b. Closed radial—(matches, several smoking objects; basis: "you need matches to light each of these").

It can be seen that the determination of the appropriate quadrant, or area, for a concept requires a cross classification, in which the effects of both variables can be taken into account with one notation. By summing the individual *S*'s scores in adjacent quadrants, two additional measures, "total public vs. private" and "total open vs. closed," can be derived. These measures provide a notation of the effect of either variable, independent of the other.

One additional measure, an "autistic index," is derived by subtracting the sum of closed-public concepts from the sum of concepts scored as open-private. While the dichotomous abstract-concrete formulation would usually lead to the indiscriminate classification of concepts within both of these areas as "concrete," a relatively high negative relationship between the two classes has been found consistently.[5]

Procedure

The object-sorting records of the schizophrenic group had been scored in a previously reported study (16). The two brain-damaged samples were drawn in the following way: All cases referred for psychological observation over a number of years for whom sorting protocols were available were drawn from the file of a large hospital facility. They numbered about 100. These test records were rescored to obtain the measures described above. The official diagnosis of each case was then obtained and the 53 cases with complete

[4] A revised set of subarea categories and supplementary instructions for scoring conceptual areas as used in this study has been deposited with the American Documentation Institute. Order Document No. 5067 from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[5] For example, with the data of the present study, the product-moment correlation between these two measures was −.48 for the brain-damaged group and −.63 for the schizophrenics.

protocols who were subsequently found to be diagnosed as brain-damaged were selected.

It was not practicable to provide as tight a control over the possibility of systematic bias in scoring as was available in the previous study (16). Therefore, several informal checks of scoring reliability were made. First, records that turned out not to be associated with a diagnosis of brain damage were compared with those of the selected sample; for each predictive measure, the average of the brain-damaged *Ss* exceeded that of the unselected group. Second, of the *Ss* who were eventually included within the brain-damaged sample, a few (about 10) were known by the scorer to have a diagnosis of brain damage because of notations on the test records. Subsequent comparison of scores for this small group against the rest of the sample indicated no systematic differences in the obtained scores. Third, the group who were subsequently identified as "brain-damaged with psychotic reaction" were found to differ considerably from the major brain-damaged sample in their obtained scores; these differences were all in a direction consistent with the hypotheses. Finally, in previous studies it has been found that with rather careful instruction and practice on scoring procedures, interscorer agreement can be raised to about 93 per cent. This compares favorably with a score-rescore agreement of 95 per cent in the present study, using 180 responses selected "blindly," and at random, from the protocols of all groups, with a time interval of about six months.

Hypotheses

In the previous study of differences in conceptualization between schizophrenic and nonpsychiatric groups, it was successfully predicted that the schizophrenic group would show "a loss of social communication (loss of 'publicness') in the formation of concepts, without evidence of a corresponding loss in high order of conceptualization (loss of 'openness')" (16, p. 45). This hypothesis grew out of a consideration of the tentative findings of one of the authors (15). In an investigation of differences in the language behavior of normal *Ss*, it was found that a predominant tendency to use closed concepts in object sorting was associated with "stimulus-boundness," or "reality-fixing"; similarly, a predominant tendency to use private concepts in object-sorting was associated with "autism." Thus, the hypothesis was based upon the assumption that "the peculiar forms of schizophrenic conceptualization should be more closely associated with autism and reality loss than with reality fixing and stimulus-boundness" (16, p. 46).

In the present study, the opposite assumption is made concerning the characteristic conceptualization of brain-damaged patients, viz., that it should be more closely associated with reality-fixing and "stimulus-boundness" (7) than with autism or loss of reality testing.

The general hypothesis of differences between brain-damaged and schizophrenic groups in object-sorting behavior is that the brain-damaged group has a greater deficit in higher order concepts (loss of "openness").

The specific hypotheses stated in terms of the operations of the study, are as follows: In comparison with the schizophrenic group, the brain-damaged group:

1. Employs significantly fewer open and significantly more closed concepts.
2. Employs significantly fewer (*a*) open-public and (*b*) open-private, and significantly more (*c*) closed-public and (*d*) closed-private concepts.
3. Scores significantly lower on an autistic index consisting of open-private minus closed-public concepts.

RESULTS

The results of an analysis of covariance for the conceptual area measures are shown in Table 2. The values shown in the "adjusted *F* ratio" column reflect only the residual variances attributable to experimental group differences after the covariance between intelligence and the conceptual area measures has been withdrawn. The extent of this covariation for each measure for *Ss* in the two groups may be seen in the column of correlation coefficients.

All six measures involved in the hypotheses indicate results consistent with the prediction of greater deficit in higher order concepts (loss of openness of conceptualization) for the brain-damaged group. Four of these group mean differences are significant at the .01 level or better.[6] The other two (open-public and closed-private) both show considerable mean differences but also show relatively high correlation values with intelligence. The greatest *F* ratio is shown for the closed-open variable, which is the major predictive measure for the present study. No hypothesis was formulated for the public-private variable, which served as the major predictive measure in the previous schizophrenic-nonpsychiatric study; in the present study, the mean difference between groups is relatively slight. It can be seen that the predominant private conceptual area for the brain-damaged group is closed-private, as contrasted with the predominance of the open-private area in the conceptual behavior of the schizophrenic group. Apparent differences in meaning for these two kinds of private conceptualization will be discussed more fully below.

A comparison of these results with those obtained in the previous schizophrenic-nonpsychiatric study (16) indicates that the extreme group separation in every conceptual

[6] *F* ratios to test for homogeneity of variance were computed for all measures in Table 2 for which *p* values for *F* ratios are shown. The results of these tests indicate that the hypothesis of a common population variance is tenable in all cases except for the open-private measure, although the ratio for the closed-private measure also approaches significance.

TABLE 2

SIGNIFICANCE OF DIFFERENCES IN CONCEPTUAL AREA SCORES BETWEEN SCHIZOPHRENIC AND BRAIN-DAMAGED GROUPS AND CORRELATION VALUES BETWEEN CONCEPTUAL AREA AND INTELLIGENCE TEST SCORES FOR THESE GROUPS

Measure	Schizophrenic $N = 37$		Brain Damaged $N = 34$		Adjusted F Ratio	p	r	r_p
	Mean	SD	Mean	SD				
Public	11.00	4.70	11.27	3.67	0.82	—	.42	<.01
Closed	6.05	3.35	11.18	3.29	25.34	<.001	−.32	<.01
Closed-Public	3.78	2.62	6.77	3.46	14.71	<.001	.04	—
Closed-Private	2.27	2.11	4.41	2.93	3.78	<.06	−.46	<.01
Open-Public	7.22	3.98	4.50	2.74	2.43	—	.48	<.01
Open-Private	5.73	3.95	3.32	2.48	10.11	<.01	−.13	—
Autistic Index*	11.95	5.95	6.55	5.13	22.99	<.001	−.29	<.02

* Open-private minus closed-public, plus 10.

area measure, except that of public-private, is between the schizophrenic and brain-damaged groups. In other words, both kinds of disorder produce departures from normal conceptual behavior, but apparently in opposite directions for at least most of the measures.

DISCUSSION

Methodological Considerations

A point relating to methodology concerns the effectiveness of cutting points for the separation of schizophrenic and brain-damaged groups. While the writers were not primarily concerned in this study with the matter of individual diagnosis, they do heartily concur with Yates (24) that such information is basic to this type of study. Using the sum of "closed" concepts as the major predictor of group differences and the total group mean (between scores 8–9) as the most reliable cutting point, the number of correct predictions for the brain-damaged group was found to be 26 out of 34 (77 per cent), and for the schizophrenics, 31 out of 37 (84 per cent). On the whole, it would seem reasonable to conclude that this type of analysis will not be very serviceable in individual diagnosis until the present techniques can be altered to provide a more sensitive measure of group differences in conceptual behavior.

Theoretical Considerations

In the writers' opinion, the Goldstein approach represented real progress in this much-neglected research area. But, more recently, a number of investigators (e.g., 3, 4, 5, 6, 9, 10, 12, 13, 15, 16, 17) have indicated that, if the venture is made to "cut through" Goldstein's dichotomous formulation of abstract and concrete conceptual levels, the yield is an improvement in prediction of group differences in conceptual behavior. The present investigators have attempted to show that the use of two isolated variables, previously embedded within the congeries of meanings for abstract-concrete, may yield significant quantitative distinctions in the conceptual behavior of clinical and "normal" groups, which are not evident when the sorting task data are analyzed within the duality of an abstract-concrete formulation. Thus, in a previous study (16), it was found that the schizophrenics showed a relative loss (or lack) of publicness in their conceptual behavior but that, in terms of order of classification, they actually showed more openness of conceptualization than did the nonpsychiatric group. In the present study, on the other hand, the brain-damaged group showed an extreme lack of openness of conceptualization; and with this measure the greatest group separation was between the two deviant groups, brain-damaged and schizophrenic.

The mean difference between the brain-damaged and schizophrenic groups on the public-private measure did not approach significance. This closeness of agreement is, in one sense, an artifact produced by canceling out opposite group tendencies on the separate conceptual area scores, open-private and closed-private; that is, in their private conceptualization, the schizophrenics are predominantly open and the brain-damaged are predominantly closed.

These differences in kinds of privateness are sufficiently pronounced to merit further consideration. The judgment of the privateness of a concept is based upon a criterion of relative inability of an observer to specify the limits of the conceptual grouping on the basis of a clearly denoted principle. Considered another way, a concept is private to the extent that the outsider would be unsure in anticipating whether additional objects would be included in or excluded from the sorter's conceptual grouping. While this criterion is applied in scoring both areas (open-private and closed-private), the quality of privateness in the two categories differs markedly.

An open-private concept has the quality of a

"personal idiom" (5). The most frequent types are idiosyncratic forms of symbolism, overabstractions, and those which show ellipses in the conceptual process. The several concepts which Cameron (3, 5) has introduced (e.g., overinclusiveness, asyndesis, metonymic distortion) to characterize schizophrenic thought are frequently apt descriptive terms for open-private conceptualization. They are so in the sense that they also refer to a form of thinking which lacks communicability, reality-testing, etc., but which involves the use of a relatively high-order of abstraction (e.g., "these are different forms of energy").

Schizophrenics, and others who employ an open-private form of thinking, frequently use some aspects of the objects in the sorting task simply to "trigger" a series of autistic associations. They are relatively insensitive to external "feedback," and they show a relative lack of self-monitoring, and lack of set toward communication. Little or none of this seems true of the thinking of those who employ predominantly closed-public concepts. Here, the quality of privateness in their conceptualization results not from *overabstraction*, but usually from *underabstraction*.

Many of the same subcategories are used in scoring both closed-private and open-private areas. The basic principle that distinguishes between these scoring areas is that of lower or higher order of conceptualization. Closed-private concepts are usually based upon the heterogeneous grouping of a number of objects "conceptualized" by the sorter at separate species-levels.[7] Open-private concepts are usually based upon a synecdochal, or other type of loose-representational, relationship perceived by the sorter among the objects within the conceptual grouping. Thus, in an example of the "closed" narrative ("I could smoke this cigar after I finish my coffee"), each of the grouped objects (cigar, matches, spoon, sugar cubes) is appropriate in its *present species form* to the "principle" of belongingness expressed. On the other hand, in an example of the "open" narrative ("fellow might want to smoke after he eats his dinner and then go bowling"), the sugar-cubes are used to represent food in general; and, even more distantly, the attribute of "greenness" of the paper square is abstracted and used synecdochally to represent the activity of bowling (inquiry revealed that the missing conceptual link was "bowling on the green").

The privateness of the "closed" narrative cited above lies in the failure of *S* to designate conceptual limits to his grouping; that is, it would be difficult for the outsider to predict whether or not each of a number of additional species-objects coming within the perceptual field of *S* would be included as "belonging" to the conceptual grouping, since no actual principle of collection is expressed. Thus, a closed-private concept is one in which there is virtually a complete *absence of a collecting principle* for a group of objects, considered by *S* to "belong together."

The virtual absence of a collecting, or generic, principle (closed-private conceptualization) and the use of conceptual classifications of a relatively low order of abstraction (closed-public conceptualization) are characteristic of the sorting behavior of the brain-damaged group. We have previously (15, 16) associated the terms "stimulus-boundness" and "reality-fixing" with this form of conceptual behavior. As distinguished from the concepts cited above to characterize schizophrenic thinking, these terms seem to us to bear a very close resemblance to some of the criteria invoked by Goldstein, Storch, Stern, and others to characterize concrete thinking in general; examples of these are impairment in discrimination between figure and ground (7), restriction of the conceptual response to the "immediate experience of the given thing or situation in its particular uniqueness" (7, p. 18), and conceptualization based upon a preponderance of concrete images (19). But it should be noted that these criteria for "concrete" behavior seem appropriate only for the specific kind of concreteness that is shown in a predominantly "closed" form of conceptualization.

On the basis of the findings of this study, the following conclusions are drawn:

1. The use of two isolated variables of conceptualization appears to be a more fruitful approach than the use of the dichotomous abstract-concrete formulation in predicting qualitative group differences in conceptualization.

2. The extreme differences in conceptual behavior between schizophrenic and brain-

[7] The term "species" is defined as "a group of objects, each of which is 'called,' or usually identified, by the same name (for example, knives, pencils)" (15, p. 185).

damaged patients suggest that the two groups should not be represented as evidencing the same type of conceptual disorder.

SUMMARY

This was a study of differences in object-sorting behavior between schizophrenic and brain-damaged groups with respect to two conceptual variables—amount of social agreement and order of conceptual classification. It was hypothesized that the brain-damaged group shows a significantly lower order of conceptual classification than does the schizophrenic group.

The variable, social agreement, was scored in terms of the relative publicness-privateness of each conceptual sorting. The second variable, order of classification, was scored in terms of the number of attributes used in the definition. In addition to these two measures, five others were derived from the interaction of the two variables.

An analysis of covariance (controlling for intelligence) indicated that all six measures for which hypotheses were formulated yielded results in the predicted direction and that four of these mean differences were significant at the .01 level, or better. The greatest F ratio was for the closed-open variable, in keeping with the major hypothesis. Differences on all of the measures, except public-private, were greatest between the schizophrenic and brain-damaged groups with the results of the nonpsychiatric group falling in between.

REFERENCES

1. BOLLES, MARJORIE. The basis of pertinence. *Arch. Psychol., N. Y.*, 1937, No. 212.
2. BOLLES, MARJORIE, & GOLDSTEIN, K. A study of the impairment of "abstract behavior" in schizophrenic patients. *Psychiat. Quart.*, 1938, **12**, 42.
3. CAMERON, N. Reasoning, regression and communication in schizophrenics. *Psychol. Monogr.*, 1938, **50**, No. 1 (Whole No. 221).
4. CAMERON, N. Deterioration and regression in schizophrenic thinking. *J. abnorm. soc. Psychol.*, 1939, **34**, 265–270.
5. CAMERON, N. Experimental analysis of schizophrenic thinking. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia*. Berkeley: Univer. of California Press, 1944, Pp. 50–64.
6. CAPPS, H. M. Vocabulary changes in mental deterioration. *Arch. Psychol., N. Y.*, 1939, No. 242.
7. GOLDSTEIN, K. Methodological approach to the study of schizophrenic thought disorder. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia*. Berkeley: Univer. of California Press, 1944. Pp. 17–40.
8. GOLDSTEIN, K., & SCHEERER, M. Abstract and concrete behavior. An experimental study with special tests. *Psychol. Monogr.*, 1941, **53**, No. 2 (Whole No. 239).
9. GRASSI, J. R. The Fairfield Block Substitution Test for measuring intellectual impairment. *Psychiat. Quart.*, 1947, **21**, 474–489.
10. GRASSI, J. R. *The Grassi Block Substitution Test for measuring organic brain pathology*. Springfield, Ill.: Charles C Thomas, 1953.
11. HANFMANN, EUGENIA, & KASANIN, J. Conceptual thinking in schizophrenia. *Nerv. ment. Dis. Monogr.*, 1942, No. 67.
12. HUNT, J. MCV. Psychological loss in paretics and schizophrenics. *Amer. J. Psychol.*, 1935, **47**, 458–463.
13. HUNT, J. MCV., & COFER, C. N. Psychological deficit. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944.
14. KASANIN, J. S. The disturbance of conceptual thinking in schizophrenia. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia*. Berkeley: Univer. of California Press, 1944. Pp. 41–49.
15. MCGAUGHRAN, L. S. Predicting language behavior from object sorting. *J. abnorm. soc. Psychol.*, 1954, **49**, 183–195.
16. MCGAUGHRAN, L. S., & MORAN, L. J. "Conceptual level" vs. "conceptual area" analysis of object-sorting behavior of schizophrenic and non-psychiatric groups. *J. abnorm. soc. Psychol.*, 1956, **52**, 43–50.
17. MORAN, L. J. Vocabulary knowledge and usage among normal and schizophrenic subjects. *Psychol. Monogr.*, 1953, **67**, No. 20. (Whole No. 370).
18. RAPAPORT, D. *Diagnostic psychological testing*. Vol. 1. Chicago: Year Book Publishers, 1945.
19. STERN, W. *General psychology*. H. D. Spoerl (Trans.). New York: MacMillan, 1938.
20. STORCH, A. The primitive archaic forms of inner experience and thought in schizophrenia. *Nerv. ment. Dis. Monogr.*, 1924, No. 36.
21. WERNER, H. *Comparative psychology of mental development*. E. B. Garside (Trans.). New York: Harper, 1940.
22. WHITE, W. A. The language of schizophrenia. *Res. Publ. Ass. Res. nerv. ment. Dis.*, 1928, **5**, 323–343.
23. VIGOTSKY, I. S. Thought in schizophrenia. *Arch. Neurol. Psychiat., Chicago*, 1934, **31**, 1063–1077.
24. YATES, A. C. The validity of some psychological tests of brain damage. *Psychol. Bull.*, 1954, **51**, 359–379.

Received February 20, 1956.

A COMPARISON BETWEEN IPSATIVE AND NORMATIVE RATINGS OF PERSONALITY[1]

JACK BLOCK

Institute of Personality Assessment and Research, University of California, Berkeley

ALTHOUGH the distinction has been noted for some years (2, 5, 6), it was Cattell (3) who first named and paid detailed attention to the difference between "ipsative" and "normative" forms of psychological measurement. In ipsative measurement, the set of scores (traits, characteristics, behavioral measures) are weighted or ordered relative to that individual's own personal mean. In normative measurement, each score for an individual is evaluated relative to the mean score of a group of individuals. An illustration of the ipsative method is the rank ordering a subject may provide of his esthetic preferences among a variety of alternatives. Here the individual has a personal norm or indifference point in terms of which his choices may be evaluated. Normative measurement is exemplified by any test which employs group norms in terms of which an individual's score achieves meaning. As Cattell remarks, the distinction has considerable implication for the propriety of various methods of statistical evaluation, for the Q vs. R controversy in factor analysis, and for diverse other issues.

One of the contexts where the ipsative-normative distinction is important is where ratings of personality are employed. For reasons of convenience, necessity, and intrinsic interest, personality evaluations by observers have been and will continue to be widely used in research. Traditionally, personality ratings have been made normatively, e.g., 100 subjects are rated as to their "aggressiveness." More recently, the Q-sort method of personality description, an ipsative procedure, has achieved some usage.

There has been some question as to the relationship existing between the two measurement approaches. Cattell suggests that ipsative measurement treated normatively will provide results different from those derived by direct normative measurement (3, p. 296). More recently, Guilford has written, "Individual differences in ipsative measurements have little meaning because there is not a single scale for all individuals . . . (Ipsative scores) should not be used for correlation of variables over a population of individuals" (4, p. 528).

Clearly, there is good logical reason for suspecting a lack of equivalence between the two measurement approaches. The fact that in the ipsative case the mean in terms of which traits are evaluated may vary from individual to individual may make possible bizarre and inappropriate relationships.

However, the matter is open to empirical investigation as well as rational analysis. The present paper reports the results of two studies in which ipsative ratings, treated normatively, were correlated with corresponding normative ratings, in a test of the functional equivalence of the two forms of measurement. The findings are, to the investigator at least, somewhat surprising. A number of current rating and assessment procedures perhaps bear revision in the light of these data.

ANALYSIS I

As part of an assessment program at the Institute of Personality Assessment and Research, 100 Air Force officers were observed for a three-day period by a staff of eight psychologists. At the end of assessment, each of the ten staff members rated each subject (*S*) on 30 aspects of personality. Each rater used a 5-point rating scale and, in evaluating the 100 *S*s, considered each rating variable singly. The ratings were afterward averaged to form for each *S* one score for each dimension. These data served as the normative measures in the comparison.

Eight of the ten staff members also described each subject by means of a 76-item Q-sort deck. The items were evaluated vis-à-vis each other for relevance and salience in characterizing each *S*'s personality, and sorted along a 9-point continuum. The scale scores for

[1] This research was supported in part by the United States Air Force under Contract No. AF 18 (600)-8, monitored by Technical Director, Detachment #7 (Officer Education Research Laboratory), Air Force Personnel and Training Research Center, Maxwell Air Force Base, Alabama. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government.

Personal views or opinions expressed or implied in this publication are not to be construed as necessarily carrying the official sanction of the Department of the Air Force or of the Air Research and Development Command.

TABLE 1

CORRELATIONS BETWEEN NORMATIVE RATINGS AND EQUIVALENT Q-SORT ITEMS: FIRST STUDY
(Reliability of normative and ipsative ratings are in parentheses)

	Ipsative-Normative Correlation	Corrected Ipsative-Normative Correlation
1. "Ability to communicate" (.90) vs. "Communicates ideas clearly and effectively" (.78)	.88	1.05
2. "Adaptive flexibility" (.87) vs. "Is rigid; inflexible in thought and action" (.80)	−.78	.94
3. "Breadth of interests" (.88) vs. "Has a narrow range of interests" (.91)	−.85	.96
4. "Constriction" (.83) vs. "Overcontrols his impulses; is inhibited; needlessly delays or denies gratification" (.88)	.82	.96
5. "Dominance" (.90) vs. "Takes an ascendant role in his relations with others" (.90)	.82	.91
6. "Evaluation of ideas" (.80) vs. "Has the ability to draw dependable and practical inferences; common sense; good judgment" (.73)	.73	.96
7. "Fairmindedness" (.81) vs. "Allows personal bias, spite or dogmatism to enter into his judgment of issues" (.69)	−.70	.93
8. "Fluency of ideas" (.90) vs. "Is stereotyped and unoriginal in his approach to problems" (.84)	−.84	.97
9. "Good judgment" (.88) vs. "Has the ability to draw dependable and practical inferences; common sense; good judgment" (.73)	.75	.94
10. "Health and vitality" (.79) vs. "Is active and vigorous" (.84)	.71	.88
11. "Impulsivity" (.83) vs. "Undercontrols his impulses; acts with insufficient thinking and deliberation; unable to delay gratification" (.75)	.69	.87
12. "Intellectual competence" (.90) vs. "Has a high degree of intellectual ability" (.82)	.86	1.00
13. "Leadership" (.94) vs. "Is an effective leader" (.80)	.79	.91
14. "Likeability" (.73) vs. "Is a likeable person (N.B. the subject's general acceptability rather than the rater's personal reactions is intended.) (.81)	.66	.86
15. "Masculinity" (.86) vs. "Is masculine in his style and manner of behavior" (.83)	.79	.94
16. "Military & social presence" (.79) vs. "Lacks social poise and presence; becomes rattled and upset in social situations" (.77)	−.79	1.01
17. "Naturalness" (.74) vs. "Is natural; free from pretense; unaffected" (.80)	.73	.95
18. "Originality" (.86) vs. "Is stereotyped and unoriginal in his approach to problems" (.84)	−.87	1.02
19. "Positive valuation of the military identity" (.79) vs. "Identifies with values, actions, attitudes, etc., of his professional group" (.72)	.62	.83
20. "Rigidity" (.79) vs. "Is rigid; inflexible in thought and action" (.80)	.77	.97
21. "Self-confidence" (.80) vs. "Lacks confidence in his own ability" (.82)	−.80	.99
22. "Self-defensiveness" (.71) vs. "Is self-defensive; rationalizes, excuses, blames" (.64)	.63	.94
23. "Self-insight" (.70) vs. "Lacks insight into his own motives and behavior" (.61)	−.66	1.02
24. "Verbal fluency" (.90) vs. "Is verbally fluent; conversationally facile" (.83)	.84	.98

each item were afterward summed across raters in order to generate for each *S* one score for each Q item. These data served as the ipsative measures in the comparison.

It proved possible to match 24 of the 30 normative ratings with Q items. Matching was done on an a priori basis by two psychologists and the criterion for matching was simply that the sensible meaning of the Q item be considered to correspond fairly closely, if not exactly, to the definition of the normative ratings.

For each of the 24 pairings, the correlation between normative and ipsative scores was computed. These are listed in Table 1, which also identifies the rating and Q-item variables, the respective split-half reliabilities, and the ipsative-normative correlations corrected for attenuation. Over all comparisons, the *median* ipsative-normative correlation, corrected for attenuation, is .95. This figure makes no allowance for the lack of exact equivalence in certain of the comparisons.

ANALYSIS II

The second comparison employed data provided by a one-day assessment of 70 applicants to medical school. A group of seven psychologists contributed normative ratings on 25 variables while a *separate* group of seven psychologists Q sorted the *S*s using a 108-item "clinical" Q deck. This deck was oriented toward more highly inferential ("genotypical") features of personality than the descriptive ("phenotypical") items used in the first study, or the normatively rated variables employed in both analyses. Because of the greater complexity of these Q items, the matching of normative ratings with ipsative items proved to be more difficult and in certain instances rather arbitrary.

TABLE 2

CORRELATIONS BETWEEN NORMATIVE RATINGS AND EQUIVALENT Q-SORT ITEMS: SECOND STUDY

(Reliabilities of normative and ipsative ratings are in parentheses)

	Ipsative-Normative Correlation	Corrected Ipsative-Normative Correlation
1. "Breadth of interests" (.83) vs. "Has a wide range of interests" (.58)	.74	1.07
2. "Dominance" (.86) vs. "Behaves in an assertive fashion" (.82)	.53	.63
3. "Ego-strength" (.84) vs. "Has a brittle ego-defense system; has a small margin of integration; precarious self-control" (.61)	−.44	−.62
4. "Impulsivity" (.75) vs. "His various needs tend toward relatively direct and uncontrolled expression; unable to delay gratification; acts out" (.59)	.64	.95
5. "Intellectual competence" (.76) vs. "Has a high degree of intellectual competence" (.69)	.55	.76
6. "Introspectiveness" (.85) vs. "Is introspective and concerned with self as an object; frequently self-aware" (.71)	.60	.77
7. "Masculinity" (.82) vs. "Is masculine in his style and manner of behavior" (.68)	.65	.86
8. "Need for achievement" (.66) vs. "Has high aspiration level for self; ambitious; wants to get ahead" (.45)	.31	.57
9. "Originality" (.69) vs. "Is original and imaginative" (N.B. *quality* of thought rather than quantity is to be emphasized) (.69)	.61	.89
10. "Responsibility" (.68) vs. "Is a dependable and responsible person" (.46)	.54	.96
11. "Self-insight" (.82) vs. "Has insight into his own motives and behavior" (.49)	.48	.75
12. "Social acuity" (.68) vs. "Is socially perceptive; responsive to a wide range of social cues" (.65)	.52	.78
13. "Social presence" (.84) vs. "Has social poise and presence; appears socially at ease" (.82)	.58	.69
14. "Verbal fluency" (.72) vs. "Is a talkative individual" (.78)	.61	.81

Fourteen matchings were deemed possible. Ipsative-normative correlations then were computed from scores derived as in the first analysis. These are presented in Table 2, along with the split-half reliabilities of the ipsative and normative measures, and the ipsative-normative correlations corrected for attenuation.

The findings from the second analysis are not quite so spectacular as in the first comparison. But the *median* ipsative-normative correlation, corrected for attenuation, is still .79. This is quite a respectable figure and is found despite the fact that the second analysis provides a much stricter and perhaps unfair test of the equivalence between ipsative and normative ratings. By using separate groups of raters in the second study, the possibility is eliminated of correlated rating error, a factor which from one point of view may have affected the results of the first analysis. But the use of separate rater groups introduces a source of variance which is extraneous to the comparison of interest, i.e., the results provided by the two methods. Because the correlation between the two measurement approaches in the second study reflects possible real differences between the two sets of raters as well as a possible real difference between the two methods, it is reasonable to find the coefficient somewhat lower.

Another cause of attenuation is to be found in the poorer ipsative-normative matchings in the second study. The ipsative ratings in the second analysis were more inferential and depended less on behavior than did the normative ratings.[8] Because the matchings were not so complete as in the first study, the ipsative-normative correlations should show some decrease.

Taking the two studies together, therefore, it would appear that there is an almost complete functional identity between the results obtained from ipsative ratings treated normatively and conventionally acquired normative rating data. The one can be directly substituted for the other, with insignificant loss of discriminations.

DISCUSSION

Why is it that the ipsative-normative ratings appear to be so fully equivalent? One possibility would seem to lie in the nature of the ipsative judgmental process. In ipsative ratings, the decision between qualitatively different variables as to "salience" is made with reference to a very large apperceptive body of information and evaluation as to the psychology of human beings. This "adaptation level" is

[8] Suggestive evidence for order of complexity as affecting the ipsative-normative correlations comes from the sets of reliability estimates of these ratings. In the first study, ipsative and normative dimensions required approximately the same level of rater inference and their reliabilities were about equal (means of .80 and .84, respectively, computed via the Z transformation). But in the second study, ipsative ratings were clearly more genotypic in their emphasis. Here the ipsative reliabilities suffered where the normative reliabilities remained the same (means of .66 and .78 respectively; computed by the Z method.)

based upon some sum or integration of the rater's lifetime experience with and conceptualization of various personalities. The frame of reference from which the ipsative judgment is made is therefore in actuality normatively derived, and it is not surprising (with after-the-fact wisdom) that the considerable equivalence exists.

But the normative reference base from which ipsative ratings are made is a much broader one than that employed for most normative ratings. In the case of ipsative ratings, the normative adaptation-level is a very stable one, accumulated over many years and not subject to radical change when a short run of disproportionate personalities is encountered. Something like an absolute scaling of the rater's impressions is obtained.

In the case of directly normative ratings, the reference frame is usually a specifically delimited population; e.g., military officers, applicants to medical school, and so on. Because in an absolute sense populations may reliably differ with respect to their mean and range on various personality variables, relationships among these variables are likely to shift appreciably as a function of the particular samples studied (1). A "masculine" graduate student may not be a "masculine" military officer when rated vis-à-vis this second population. But in the generalized normative frame in terms of which the ipsative rating is made, his rating for "masculinity" is likely to remain the same, regardless of the particular sample of which he is a member. Ipsative ratings, normatively treated, thus offer the possibility of providing more stable interrelationships among variables than has usually been the case.

Which method, ipsative or normative, is to be preferred in research settings? The normative method clearly has a more immediate convenience since only one variable is dealt with at one time and the subject sample may be readily evaluated in terms of the particular one or two dimensions of interest. To achieve ipsative scores on but a few dimensions, *all* of the variables in the ipsative sample of traits must be evaluated for *each* subject, a task which may be quite formidable. It should be noted, though, that the normative approach loses its advantage of convenience as the number of variables of interest increases.

Although there are many contexts where the ipsative method is inefficient or inappropriate, there is one somewhat unusual benefit to be enjoyed from its use. Frequently, in research retrospect, e.g., in the summing up of longitudinal studies, one wishes that an additional variable or conception of personality had been evaluated with reference to the now departed subject sample. With the normative method, there is no simple or easily justified way in which these belated evaluations may be gathered. Memories prove faulty or the originally participating raters may no longer be available. With ipsative data, however, the desired ratings are quickly achieved.

As an empirical illustration of how the omission may be rectified, consider the variable "Over-all Military Effectiveness in a Command Position," normatively rated in the first study reported above. This variable could not be matched with any of the 76 ipsative Q items and was not employed in the main body of the analyses. Let us suppose that some years after the actual study, some evaluation of this dimension was desired for criterion purposes. An ipsative measure could be obtained in the following manner:

1. Each of a number of staff psychologists would Q-sort their conception of the Air Force officer manifesting the greatest degree of "Over-all Military Effectiveness in a Command Position." In actual fact, the average intercorrelation of five conceptual Q sorts was .80. Only two of the five sorters of the concept had previously observed the officer sample and their sorts did not interrelate differently than did the remaining three.

2. A composite of this conceptual model would then be constructed by summing the Q-item placements across sorters. In the actual case, this composite has a reliability of .89.

3. The empirical Q sorts of each of the 100 officers would then be correlated with the conceptual composite.

4. These correlations with the composite could be treated as a set of scores ordering the subjects on the specified dimension. In the actual illustration, when these correlations, considered as scores, were correlated with the initially obtained normative ratings of the dimension, the uncorrected correlation was .85. Substantially the same degree of equivalence existed in several other tests of this technique.

Thus it would appear that by use of conceptual Q sorts, and an original Q set which is comprehensive in nature, scores can be derived from actual Q sorts which pertain to dimensions not specifically included in the Q set or otherwise anticipated. The normative method does not offer this useful facility in any convenient form, in part because normative ratings are usually too few in number, but most particularly because normative ratings as usually gathered are too tied to the particular referent population.

It is perhaps wise to close on a cautionary note. The present paper deals only with the comparison of ipsative and normative *ratings*. The ipsative-normative distinction has ramifications much beyond those touched upon here. In particular, the issues involved in comparing ipsative and normative *objective measures* await empirical study.

SUMMARY

Ipsative and normative ratings of personality were compared in two separate studies. Both analyses showed an almost complete equivalence between the two methods. A possible reason for this relationship was discussed and some of the advantages of the ipsative approach presented.

REFERENCES

1. BLOCK, J. The difference between Q and R. *Psychol. Rev.*, 1955, **62**, 356–358.
2. BURT, C. Correlations between persons. *Brit. J. Psychol.*, 1937, **28**, 59–96.
3. CATTELL, R. B. Psychological measurement: ipsative, normative, and interactive. *Psychol. Rev.*, 1944, **51**, 292–303.
4. GUILFORD, J. P. *Psychometric methods*. New York: McGraw-Hill, 1954.
5. STEPHENSON, W. Correlating persons instead of tests. *Charact. & Pers.*, 1935, **4**, 17–24.
6. THOMSON, G. *The factorial analysis of human ability*. London: Univer. of London Press, 1939.

Received February 21, 1956.

THE EFFECTS OF GROUP EXPECTATIONS AND SELF-ESTEEM UPON SELF-EVALUATION[1]

EZRA STOTLAND, STANLEY THORLEY, EDWIN THOMAS, ARTHUR R. COHEN,[2] AND ALVIN ZANDER

University of Michigan

RESEARCH on the level of aspiration has demonstrated quite well that a person's evaluation of his performance does not depend upon his absolute level of achievement alone. What counts is the level of achievement relative to his aspiration—his goal line. If the achievement lies on or above the goal line he will probably rate his performance as good, depending upon the size of this difference. These studies have been carefully reviewed and interpreted by Lewin *et al.* (5).

The placement of the goal line, whether it is high or low, may be influenced by different scales of reference. Two kinds of reference scales have been mentioned often. One is a source of pressure from other persons or groups to perform in a particular fashion on that activity. The other involves pressure from the person's own expectations concerning his likely performance in that activity, arising from past experiences. In the present study we examine the effect of a specific level of achievement upon an individual's evaluations of his performance when the achievement is relative to a level of aspiration established by a group, and when the achievement is also relative to the member's stabilized expectations about himself as represented by his self-esteem.

Group Expectations

The ability of a group to determine levels of aspiration for individuals is discussed by Lewin *et al.* (5). Several different approaches have been used in studying these events. In the method most close to the purposes of this study, a group's level of aspiration was described to the members and they then accepted this level as their own (2).

This approach suggests that we may conceive of the level of aspiration as a group standard, which is the resultant of pressures to achieve that level exerted by the group upon the members. The power of a group is its ability to influence the beliefs or behavior of its members. If the group has enough power to influence the members, and the group standard is high, the members will set high levels of aspiration for themselves. If the standard is low, they will set low levels of personal aspiration (6).

The power of a group to influence its members may vary under different conditions. It is evident in the work of Festinger (3) that a group's power is greater the more that the group is attractive to its members. Thus, the members' levels of aspiration will be more closely modeled after the aspiration of the group, the more that the group is attractive to them. If attraction to the group is constant and moderately high among all members, however, another matter determines whether they conform to their group's level of aspiration. This is the relevance of the activity.

Relevance refers to the degree to which an activity is important to the purposes of, or for the maintenance of, a group. Schachter (7) has shown that the power of the group is likely to be stronger upon persons engaging in an activity relevant to the group than upon persons performing a nonrelevant activity. It follows, that the aspiration level set by a group concerning a relevant task tends to be supported by stronger pressures on the person to perform up to that level, than does an aspiration level set for a nonrelevant task.

Regardless of the relevance of an activity, some persons may accept a group aspiration as their own more than do other individuals. We will not occupy ourselves here with the reasons for variation in accepting the group's level of aspiration. It is expected, however, that the group's level of aspiration would have more effect upon personal evaluation of a success or a failure, the more the members accept the level of aspiration as their own.

For many groups in our society there are

[1] The work on this study was financed by a grant-in-aid [M-701 (C)] from the National Institute of Mental Health of the National Institutes of Health, Public Health Service, Alvin Zander, principal investigator.

[2] Arthur R. Cohen is now at Yale University.

external criteria which help a member to judge how well he has performed an activity which is of interest to the group. An athlete knows whether he has won, and a salesman knows whether he has made the sale. The team's, or staff's, level of aspiration is not needed for matters such as these to help a person know if he has succeeded or failed, but the group's aspiration is pertinent in influencing the extent to which the individual evaluates an objective failure as a failure and an objective success as a success, assuming, of course, he expects that the group will become aware of his performance. In the present study the subjects were made either to fail or to succeed, and had a clear knowledge of the level of their performance. The major issue is: what conditions in the group may influence a person's evaluation of his performance when he knows that he has succeeded or failed?

From the considerations thus far, two general propositions are derived: (*a*) When a task is relevant to a group, the group's level of aspiration will raise or lower a member's evaluation of his performance more, after an experience of objective success or failure, than when the task is nonrelevant to the group. (*b*) The group's level of aspiration will raise or lower a member's evaluation of his performance to the degree that he accepts the group's level of aspiration as his own.

Specific hypotheses suggested by these propositions, but more amenable to verification, are as follows:

1. A level of aspiration set by a group for its members is more often accepted by a member as a personal goal when the activity is relevant to the group than when it is nonrelevant.

2. A failure on a relevant task generates poorer self-evaluation of a person's performance if a group's aspirations are high than if they are low (while the evaluation of a failure on a nonrelevant task is not affected by different levels of the group's aspirations).

2-a. A failure when a person accepts the group's level of aspiration as his own, generates lower self-evaluation of his performance if the group's aspirations are high than if they are low (while evaluation of a failure, when the group's aspiration is rejected, is not affected by different levels of the group's aspirations).

3. A success on a relevant task generates better self-evaluation of a person's performance if a group's aspirations are high than if these aspirations are low (while the evaluation of a success on a nonrelevant task is not affected by different levels of the group's aspirations).

3-a. A success when a person accepts the group's level of aspiration as his own, generates better self-evaluation of his performance if the group's aspirations are high than if they are low (while the evaluation of a success when the group's aspiration is rejected is not affected by different levels of the group's aspirations).

Self-Esteem

When a person evaluates his level of achievement he refers not only to the expectations of other persons but to the concept he has of himself as well. The degree of an individual's self-esteem, therefore, is expected to affect the way in which an individual evaluates his performance in a particular situation. *Self-esteem* is a function of the degree of coincidence between an individual's ideal and actual concepts of himself. The more that the actual and ideal concepts coincide, the higher the self-esteem.

Persons who are high in self-esteem, we assume, differ from those who are low in the way they react to an immediate experience which has been a test of their abilities. Cohen (1) has reported that persons with high self-esteem, as compared to those with low, are characterized by tendencies to protect themselves from negative self-evaluation. It is not clear why this should occur. Perhaps persons with high self-esteem have learned to overlook their objective failures and to concentrate upon their successes in building their concept of themselves, or perhaps they have consistently performed in accordance with their ideals and hence are less threatened by an experience of failure. In any event, it would be expected that persons with high self-esteem would be able to evaluate an objective failure as a small failure and an objective success as a large success. Persons with low self-esteem, however, because they do not protect themselves from negative evaluation, are more likely to evaluate an objective failure as a very poor performance and a success as a small success.

When the person is a member of a group which holds certain standards for his performance, how does the level of his self-esteem affect his evaluation of his achievement?

A number of different answers to this question appear plausible. A choice among the alternatives cannot be made with confidence, however, because there is no way of anticipating whether the potency of a scale of reference provided by the power of the group will equal that offered by a person's self-esteem; or whether the weight of the group and individual influences will differ under certain conditions, but not under others. Speculation about these matters is postponed until it is clear how the group standards and self-esteem interact in determining the way a person evaluates his level of achievement.

METHOD

An opportunity to examine these phenomena was created in a laboratory experiment in which the effects of both group pressures and self-esteem upon self-evaluation could be observed. Two levels of group aspiration were created, termed high and low expectations hereafter. At the same time, the relevance of the activity to the group's purpose was varied by describing the *S*'s task as essential to the group goal in one set of conditions and nonrelevant to the group goal in the other. Thus, there were four conditions consisting of high or low expectations in combination with relevant or nonrelevant tasks. Finally, within each of these conditions, half of the *S*s were made to fail on their tasks and half to succeed. The dependent variables were measured in a questionnaire administered after the *S*s had worked on the tasks but while they were assuming that the experiment was not yet finished. The variables were the *S*s' evaluations of their performances on the task and their concern with the group's expectations (i.e., acceptance of the group's level of expectations). All were measured on 6-point graphic rating scales. In addition, a measure of self-esteem was administered before the experiment was started, and again as part of the second questionnaire.

Measurement of self-esteem. The measure of self-esteem was designed in accord with the definition of the concept and modeled after an instrument used by Cohen (1). A person's score describes the degree to which he meets his needs in ways he considers to be ideal.

The items of the test area series of brief paragraphs each describing a hypothetical situation in which a person faces a potential frustration of a need. Problems of meeting needs in five representative areas are included: achievement, recognition, affiliation, cognition, and autonomy. After reading each paragraph, the *S* indicates how a person would react by choosing from among five behavioral alternatives. He first chooses the reactions *most* and *least* descriptive of the way a person should ideally behave. Then, after a second reading of each paragraph, he chooses the reactions *most* and *least* descriptive of the way a person actually would behave.

There follows an example of one of the hypothetical situations and the alternative responses for it.

A person has always valued his sense of freedom and independence. Because of what others think and say, this sense of freedom and independence is greatly hampered.

	Most	Least
a. Reason that maybe a great many mutual responsibilities with others are most important after all.	•	•
b. Figure that while things are confining now, knowing where one stands in relation to others is half the battle.	•	•
c. Tend to turn his thoughts to more pleasant aspects of his dealings with others.	•	•
d. Face the fact that other people are over-eager to maintain their opinions and ideas in the face of his.	•	•
e. Think twice before overly involving himself with others because these experiences are uncomfortable for him.	•	•

The responses are scored by giving a +1 for each occasion where the actual most preferred reaction is the same as the ideal most preferred reaction, and where the actual least preferred is the same as the least preferred. A score of −1 is given for a coincidence of the actual most preferred with the ideal least preferred, and of the actual least preferred reaction with the ideal most preferred. If no coincidence occurs, no score is given. The person's score is obtained by summing across all items.

In previous work with a longer version of this measure on a sample of sixty persons, the odd-even split-half correlation was .84 (.91 when corrected by the Spearman-Brown formula). In order to measure self-esteem in the laboratory, however, the requirements made it necessary to use only one-third of the hypothetical situations of the original instrument. Scores on all the items were significantly related to total scores in the larger instrument.

The sequential phases of the experiment are now described.

Group formation on a cooperative task. The purposes of the opening part of the experiment were, first, to form a basis for the subsequent manipulation of relevance and, second, to make membership in the group salient for the *S*s. Making membership salient, it was assumed, made attraction to the group high so that group pressures were acceptable to the members. These purposes were accomplished by having the group work and succeed on a cooperative task, in competition with other, fictitious groups.

The *S*s were 175 male undergraduates at the University of Michigan who volunteered to participate in the study. They came to the laboratory four, five, or six at each session, and care was used to ensure that they were strangers to one another. As they came into the laboratory, they were taken to individual cubicles to fill out the self-esteem questionnaire. When all of them had finished, they were seated in a circle to receive instructions, and were told that they would be working

on a cooperative task in competition with other teams. In order to help create a "team spirit," they were asked to pick a name for themselves.

The group task given them was to form two letters of the alphabet. The letter *U* was first to be formed in a groove in a board by filling it with various pieces of different sizes and shapes, as in a jigsaw puzzle. After each *S* had been given 12 pieces, the board was passed from person to person around a circular table 12 times. Each man placed one piece in the groove each turn, flush against the pieces already there. Thus, each person made an equal contribution to achievement of the group goal: to complete the *U* as fast as possible.

When the group was finished, the *E* announced the time and publicly placed a pin for the team on a graph on which the speeds of the "other" groups were indicated. This graph was arranged so that each group discovered it had done moderately well relative to others. Moderate achievement was stressed in order to prevent strong feelings of either success or failure which could confound the effect of later procedures.

Relevance and nonrelevance inductions. After the group had completed its *U*, the *S*s were given tasks to be done in private, with different instructions for the relevant and nonrelevant conditions. When the individual task was made relevant, the *E* told them that the second letter would later be formed by the group from parts each member would provide. The one piece in his individual puzzle which fitted within the design, but did not touch any of the sides, was the piece required by the group. The *S*s were then told to return to their cubicles to work on the individual puzzles. They were allowed 12 minutes for the task.

In the nonrelevant condition, the *S*s were asked to work on the same task, but were told that it had nothing at all to do with the group goal and that they would each receive individual scores on whether or not they were able to complete the task. They were informed that they would return to the second group letter at a later time.

Expectations induction. The two levels of expectations, high and low, were induced as follows. After explaining the individual tasks, and before the *S*s began to work on them the *E* said:

"We often find it helpful in group situations like this to give each person a clear idea of how well his group expects him to perform. It might be that the group expects him to meet a given level of performance, or it might be that the group feels it cannot really expect a member to do that well on a difficult task. I would like you to decide which of these expectations you want to impose on yourselves. I am going to pass out these ballots on which you can indicate which of these expectations you want. Please read the descriptions of the expectations very carefully before voting. I will collect and count the ballots."

The questions read as follows:

Please check one

1. *High expectations:* I vote that each of us will be expected by the others to complete his individual task, and that we will expect the others to do the same.

2. *Low expectations:* I vote that each of us will be expected to do only as well as he can toward completing his individual task, and that we expect the others to do the same.

The two expectations conditions were created by giving the *S*s contrived tabulations favoring high or low levels.

Achievement level. In the success condition, the task was easy enough to allow 95 per cent of the *S*s to complete it in the allotted time. In the failure condition, the *S*s were given puzzles which were impossible to complete because no piece met the criterion of a successful solution. Blank jigsaw puzzles (i.e., no pictures on the faces of the pieces) were used in order to make it uncertain as to how close they had come to the correct way of assembling the parts if they did not complete it. Thus a failure could only be an objective failure in fact, and no gradations of partial success were possible.

Questionnaires. After the *S*s were through working on the individual task, they were again given a questionnaire with the understanding that they would do the second group letter after the form was finished. When the questionnaire was completed, they were told that there was no second group task and the experiment was over. The purpose of the study was then explained, and all questions answered.

RESULTS[3]

Group expectations and self-evaluation. The first hypothesis, which asserts that members will accept the level of the group's expectations as a personal goal line more often when the task is relevant than when it is nonrelevant, was tested by examining the degree to which the *S*s said they were "concerned" about the team's expectations while they had been working on their individual tasks. The mean concern in the relevant condition is 4.00, which is significantly greater than the mean of 3.27 in the nonrelevant condition ($p = .01$, by t test). The data support the hypothesis.

The second hypothesis states that a person who fails on a task which is relevant to the group will evaluate his performance lower, if the group's expectations are high than if the group's expectations are low. In Table 1 it may be seen that in the relevant condition the mean self-evaluation is lower (2.05) when the group expectations are high, than when the expectations are low (2.65). The difference between these means is significant by a one-tailed t test at the .05 level. When the task was nonrelevant, and the members failed, the

[3] A comparison of any two means which is not specifically mentioned in the text may be assumed by the reader to be not significant at the .05 level or better, except for the contrasts between means of self-evaluation in success conditions with those in failure conditions, all of which are significant.

level of the group's expectations creates no significant difference between the mean evaluations of performance (2.00 and 1.90). The data support the second hypothesis.

Hypothesis 2-a states that when persons fail on an activity, they will evaluate their achievement as poorer when the group's expectations are high, and will evaluate their performance as better when the group expectations are low. This effect will be more marked for persons with high concern than for those with little concern about the group's expectations.

In Table 2 the mean self-evaluation for those *S*s with high concern for the group's expectations is significantly lower, where the group's expectations are high, for a relevant task which is failed (1.83), than where the group's expectations are low (2.83). This difference is significant by a *t* test at .05. The persons who were little concerned about the group's expectations do not differ significantly in their evaluation of their failure experience according to whether the group's expectations were high or low (2.36 and 2.20 respectively). When the task was nonrelevant, the members' degree of concern about the group's expectations does not generate a significant difference in the way that they evaluate a failure. It appears that the group's level of expectations had its strongest effect upon self-ratings of failure when the task was relevant and the members were highly concerned about the group's expectations. These results support Hypothesis 2-a.

The third hypothesis proposes that a person who succeeds on a relevant task will evaluate his performance lower when group expectations are low, and higher when group expectations are high. No significant differences are found (Table 1) in comparing the means of those who

TABLE 1

MEAN SELF-EVALUATION OF SUCCESS ON TASK FOR ALL CONDITIONS

(*N*'s in parentheses)

Achievement level	Relevant task		Nonrelevant task	
	High group expectations	Low group expectations	High group expectations	Low group expectations
Failure	2.05 (25)	2.65 (21)	2.00 (20)	1.90 (25)
Success	4.80 (22)	4.59 (22)	4.95 (20)	4.80 (20)

TABLE 2

MEAN SELF-EVALUATION OF SUCCESS ON TASK FOR ALL CONDITIONS, COMPARING PERSONS WITH HIGH AND LOW ACCEPTANCE OF GROUP'S EXPECTATIONS

(*N*'s in parentheses)

Achievement level	Relevant task		Nonrelevant task	
	High group expectations	Low group expectations	High group expectations	Low group expectations
A. High acceptance of group's expectations				
Failure	1.83 (12)	2.83 (12)	1.80 (10)	2.40 (5)
Success	4.73 (15)	4.69 (13)	5.09 (11)	5.00 (9)
B. Low acceptance of group's expectations				
Failure	2.36 (11)	2.20 (10)	2.60 (10)	1.69 (16)
Success	4.50 (10)	4.60 (10)	4.78 (9)	4.69 (16)

succeeded, regardless of the group's level of aspiration or the relevance of the task. Thus, the third hypothesis is not supported. In Hypothesis 3-a it was predicted that a self-rating of success after a success experience would be a direct function of the group's level of expectations, more so for persons with high concern than for those with little concern. The anticipated results did not occur; thus Hypothesis 3-a is not supported.

Self-esteem and self-evaluation. We examine now the effect of an individual's self-esteem upon his evaluation of his achievement. Self-esteem was considered to be an enduring personality characteristic which the *S*s brought with them to the laboratory. It was assumed that it would not be changed by the experience there. The reasonableness of this assumption is shown by a comparison of the mean self-esteem scores obtained prior to the experiment with scores obtained at the time of the final questionnaire. No significant differences appeared between any of the experimental conditions. Thus, the events in the experiment did not change the participant's self-esteem to a significant degree.

Persons with high self-esteem have little more positive evaluation of their own performance than those with low self-esteem, when all other matters are equal. The mean self-ratings made by persons with high self-esteem is 3.95 and the mean for those with low self-esteem is 3.88.

Members who failed on the task, however, who were high in self-esteem evaluated their

TABLE 3

MEAN SELF-EVALUATION OF SUCCESS ON TASK FOR ALL CONDITIONS, COMPARING PERSONS WITH HIGH AND LOW SELF-ESTEEM

(*N*'s in parentheses)

Achievement level	Relevant task		Nonrelevant task	
	High group expectations	Low group expectations	High group expectations	Low group expectations
A. Persons with high self-esteem				
Failure	2.36 (8)	2.60 (15)	2.67 (12)	2.45 (11)
Success	4.70 (10)	4.42 (12)	4.82 (11)	5.25 (8)
B. Persons with low self-esteem				
Failure	1.86 (14)	2.43 (7)	1.50 (8)	1.20 (10)
Success	4.87 (15)	4.78 (9)	5.11 (9)	4.59 (17)

performance higher than those with low self-esteem, all other variables constant. The means are 2.55 and 1.75 respectively, and are significantly different at the .05 level. Among the persons who succeeded, those who were high in self-esteem were not at all different from persons low in self-esteem in the self-ratings they assigned to their achievement (4.75 and 4.83 respectively). The amount of the member's self-esteem, then, affected evaluation of performance when the person failed, but not when he succeeded.

It was apparent earlier that the members were more likely to evaluate a failure in accordance with the group's level of expectations when the task was relevant than when it was nonrelevant. It is striking, now, to observe that self-esteem appears to have served as a scheme of reference for the members more in the nonrelevant condition than in the relevant one. In Table 3 it may be seen that where the expectations of the group are *high* and the person fails on a task which is nonrelevant, those with high self-esteem have higher evaluations (2.67), than the persons with low self-esteem (1.50). The difference between these means is significant at the .05 level. Where the expectations of the group are *low* and the persons fail on a task which is nonrelevant, those with high self-esteem again have higher self-evaluation (2.45) than the individuals with low self-esteem (1.20). The difference between these means is significant at the .01 level. When the *S*s failed on a relevant task, no significant differences were found in the self-ratings of those with high self-esteem compared to those with low self-esteem. It appears that the members were more likely to evaluate their performance according to their levels of self-esteem when the power of the group was weakest (i.e., the task was nonrelevant), but less likely to do so when the power was strongest (i.e., the task was relevant). This statement is true, however, only when the members failed. Among those who succeeded on the task, persons with high self-esteem had no different feelings of success than those with low self-esteem when level of expectation was held constant, regardless of the relevance of the task.

In the opening discussion it was anticipated that persons with high self-esteem would better protect themselves against poor evaluations than those with low self-esteem. We have noted that persons with high self-esteem compared to ones with low self-esteem, evaluate their performances more favorably when they have failed. We have also noted that when self-esteem alone is varied there is no difference in the evaluation of their performance. These last data suggest that persons with high self-esteem do not always rate their performance more highly than persons with low self-esteem. Can we say, then, that persons with high self-esteem appear to protect themselves? Do they do this regardless of the group's expectations? The data in Table 4 are pertinent to these questions. They indicate that persons with high self-esteem, compared to individuals with low self-esteem, are cautious about becoming concerned with the group's expectations.

The degree of concern felt for the group's

TABLE 4

MEAN CONCERN WITH GROUP EXPECTATIONS FOR ALL CONDITIONS, COMPARING PERSONS WITH HIGH AND LOW SELF-ESTEEM

(*N*'s in Parentheses)

Achievement level	Relevant task		Nonrelevant task	
	High group expectations	Low group expectations	High group expectations	Low group expectations
A. Persons with high self-esteem				
Failure	4.25 (8)	4.20 (15)	2.83 (12)	2.91 (11)
Success	3.40 (10)	3.72 (13)	3.18 (11)	4.11 (9)
B. Persons with low self-esteem				
Failure	3.86 (14)	3.69 (7)	4.63 (8)	2.70 (10)
Success	4.67 (15)	3.63 (8)	3.67 (9)	2.94 (17)

expectations differed for persons with high self-esteem compared to those with low esteem, on the nonrelevant tasks. As may be seen in Table 4, persons with high self-esteem stated that they had little concern for the group's expectations when they had failed and the group's expectations were high (2.83). Those with low self-esteem, on the other hand, when they had failed under high group expectations, expressed high concern for the group's expectations (4.63). The difference between these means is significant at the .01 level. In addition, participants with high self-esteem said they were most concerned about the group's expectations when they had succeeded and the level of the group's expectations was low (4.11). While *Ss* with low self-esteem were little concerned with the group's expectations when they had succeeded and the group's expectations were low (2.94). The difference between these means is significant, however, at only the .10 level.

It appears that *Ss* with high self-esteem became interested in the group's expectations when the expectation levels were easy to achieve and in fact had been reached, and were less interested in the group's expectations when these expectations were difficult to reach and they had failed to perform up to that level. Persons with low self-esteem were in contrast: they became attentive to the group's wishes when these were difficult to gain and the person had failed, while they were less attentive to the expectations when the level was easily available and they had already succeeded. The results suggest that persons with high self-esteem better protect themselves against unfavorable evaluation by avoiding involvement in the group's expectations when an unfavorable comparison with others would be likely to occur.

In the relevant situation the member's self-esteem was not related to his degree of concern over the group's expectations to a significant degree.

DISCUSSION

It is to be remembered that the *Ss* had the criterion for a successful performance clearly in mind. They knew whether they had succeeded or failed on their individual tasks at the time they were asked to stop. With this in mind we may wonder why the *Ss* should pay attention to the expectations of their group. Festinger (4) has offered a proposition which points to the salience of this question: "When an objective, nonsocial basis for evaluation of one's ability or opinion is readily available, persons will not evaluate their opinions or abilities by comparison with others." It is clear from the context of this quotation that Festinger means to include conformity to group pressures, as well as simple comparison of own performance with the achievement of others.

The results suggest that persons who were successful did not tend to evaluate their performance in the light of the wishes of the group, but that those who failed were affected by the group's expectations. Successful persons clearly evaluated their performance as highly successful regardless of whether the task was relevant or nonrelevant, the group's expectations were high or low, or the group's expectations were an important concern to them or not. Failing persons, however, evaluated their performance low if the group had high expectations, and they evaluated their performance high if the group had low expectations. The effect of the level of expectations upon the evaluation of their performance was also greater the more the persons were concerned about the expectations of the group. These results occurred when the task was relevant, but did not occur when the task was nonrelevant.

Why did the wishes of the group determine self-evaluations for persons who failed more than for ones who succeeded? A failure experience, we believe, has an important property which is not present in a success. The person who has failed cannot be sure how the group will receive his report that he has failed. An audience for a report of failure, especially when it is an attractive group, is likely to create some embarrassment for the reporter. Uncertainty concerning the consequences of a report of failure makes the person who has failed highly sensitive to the opinions and evaluations of the others. Thus, social reality is more important to a person who fails than to a person who succeeds. The more the relevance of the task, the more does the person who has failed feel impelled to evaluate himself in accordance with the group's expectations. The person who has succeeded, in contrast, can be quite sure that his report of success will be

favorably received regardless of the group's level of expectations. He does not need the group for social reality and is quite ready to evaluate a success as a success regardless of the group's wishes in the matter. These results and their interpretation suggest that even though an objective, nonsocial basis for self-evaluation is available, a person may evaluate himself by comparison with others, if he has failed on a task which is relevant to the group.

It is noteworthy that relevance or nonrelevance of the task, exclusive of the group's expectations, did not affect the evaluation that a person placed upon a failure experience. These results indicate, we believe, that "letting a group down" on a relevant task is no more conducive to poor self-evaluation than failure on a nonrelevant activity. The evaluation a person assigns to his performance is primarily affected by the group's level of expectations, where the task is relevant.

The results concerning the effect of self-esteem suggest several interesting issues. Why, for example, did self-esteem affect the member's evaluation of his achievement more when he had failed on a nonrelevant task than when he had failed on a relevant one? If it is assumed, as we have done, that group pressures are less potent in the nonrelevant condition, it is probable that the influence of self-esteem upon personal evaluation was relatively stronger than the power of the group in that condition. In the relevant condition, however, both because relevance increases the power of the group, and because a failure experience made the group more potent as a scale of reference, self-esteem was relatively weaker than the group as a determinant of the self-evaluation. This belief has been supported by the findings from several studies, as yet unpublished (1, 8). In these investigations, variation in the behavior of *Ss*, which was predicted because of variation in their personality characteristics, occurred more often when influences from the group were weak than when influences from the group were strong.

The results indicate that persons with high self-esteem, as we have conceived of it and measured it, dealt with their experiences in the laboratory in a way that helped them to maintain their high self-esteem. They tended to evaluate themselves better after a failure than individuals with low self-esteem, regardless of the level of the group's expectations. Furthermore, they became more concerned about the group's expectations when favorable evaluation was readily possible for them. Persons with low self-esteem reacted to their experiences in a way that made it difficult to improve their self-esteem. They not only evaluated a failure very low, but they became more concerned with the group's expectations when an unfavorable evaluation was most likely for them. Persons with high self-esteem, in brief, reacted to their experiences in the group in such a way that they best protected themselves from unfavorable evaluation.

SUMMARY

In a laboratory experiment, members were made to believe that their group had either high or low expectations for the quality of their achievement on individual tasks. The tasks were described as either relevant or nonrelevant to the purposes of the group, thus varying the power of the group over the member's activity. Half of the *Ss* within each of these four conditions were allowed to succeed and the other half were made to fail. In addition, measures of self-esteem were available for all participants.

The following hypotheses were supported.

1. A level of aspiration set by a group for its members is more often accepted by a member as a personal goal when the activity is relevant to the group than when it is nonrelevant.

2. A failure on a relevant task generates poorer self-evaluation of a person's performance if a group's aspirations are high than if they are low (while the evaluation of a failure on a nonrelevant task is not affected by different levels of the group's aspirations).

2-*a*. A failure, when the person accepts the group's level of aspiration as his own, generates lower evaluation of his performance if the group's aspirations are high than if they are low (while evaluation of failure when the group's aspiration is rejected is not affected by different levels of the group's aspirations).

Hypotheses concerning the effect of the level of the group's aspirations upon the evaluation of a successful performance were not supported. It was proposed that a person is more likely to evaluate a failure than a success

in reference to the group's expectations; because the failure makes him more dependent upon the group for social reality concerning evaluation, than does success.

Persons high in self-esteem protected themselves against unfavorable evaluations better than those low in self-esteem.

The group's expectations appear to have been more potent as a scale of reference than the individual's self-esteem in determining his evaluation of his performance. When the influence of the group was weakest (task was nonrelevant) persons high in self-esteem, compared to those who were low, differed in the way they evaluated their performance. When the influence of the group was strongest (task was relevant) there was no difference in the way that persons high or low in self-esteem rated their achievement.

REFERENCES

1. COHEN, ARTHUR R. The effects of individual self-esteem and situational structure on threat oriented reactions to power. Unpublished doctor's dissertation, Univer. of Michigan, 1953.
2. HERTZMAN, M., & FESTINGER, L. Shifts in explicit goals in a level of aspiration experiment. *J. exp. Psychol.*, 1940, **27**, 439–452.
3. FESTINGER, L. Informal social communication. *Psychol. Rev.*, 1950, **57**, 271–282.
4. FESTINGER, L. A theory of social comparison processes. *Hum. Relat.*, 1954, **7**, 117–140.
5. LEWIN, K., DEMBO, TAMARA, FESTINGER, L., & SEARS, PAULINE S. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944.
6. RASMUSSEN, G., & ZANDER, A. Group membership and self-evaluation. *Hum. Relat.* 1954, **7**, 239–252.
7. SCHACHTER, S. Deviation, rejection, and communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 190–207.
8. THOMAS, E. Effects of role interdependence and ego strength on group functioning. Unpublished doctor's dissertation, Univer. of Michigan, 1956.

Received February 29, 1956

AGE DIFFERENCES IN VERBAL LEARNING[1]

SHELDON J. KORCHIN AND HAROLD BASOWITZ

Institute for Psychosomatic and Psychiatric Research and Training, Michael Reese Hospital

THE aphorism that "you can't teach an old dog new tricks" undoubtedly predates psychological research both in the field of aging and of learning. However, the specific conditions which determine the relative ease or difficulty of learning for the older person, the importance of different types of material, learning situations and the like, have been a concern of systematic and experimental work within the psychology of learning only in relatively recent years.

Of considerable interest is a 1934 study by F. L. Ruch (13) which has since become a basic reference in discussions of age-related changes in human learning. With the view that senility can be conceived as a condition of lowered biological plasticity, Ruch hypothesized that older compared to younger persons should show a greater deficit in any learning situation which requires the reorganization of existing habits, less deficit where earlier experiences can be utilized in new learning, and have an intermediate amount of deficit where old habits neither aid nor block new learning.

In his experiments Ruch (13) had three age groups learn a pursuit rotor task, first in direct vision, then in mirror vision, and three lists of paired-associate materials. These lists differed in the degree to which they utilized or contradicted earlier learning: (*a*) paired associates of familiar words, logically related; (*b*) nonsense materials (arbitrary equations of the form $E \times Z = G$); and (*c*) interference materials (false equations of the form $3 \times 5 = 6$). These tasks were presumed to differ not only in difficulty and novelty but also in the degree to which prior experience might enter to facilitate or disrupt present learning. In the word-associate task earlier learning should facilitate the learning of these simple associates; the nonsense equations required the learning of essentially new material which should be unaffected by earlier information; while in the case of the false equations the material was not only new but contradictory to previously-formed familiar associations. With such material one might expect competing responses, the highly overlearned true products of simple multiplications, to interfere with the acquisition of the new habit.

Comparing the differences in mean scores between age groups on each of the procedures, Ruch (13) found that the young-old difference was greater for the mirror than direct vision perceptual-motor task and, among the verbal tests, was greatest for the interference material, less for the nonsense task and least for the familiar word associates. On this basis Ruch concludes that older persons have least "deficit" in learning materials which are compatible with habitual behavior and greatest where the new learning contradicts earlier habits. However, if one compares the scores for each procedure within each group separately, it is interesting to note that all groups performed *better* on the interference material than on the nonsense task, although the *difference* between young and old is greater on the former than on the latter.

As part of a larger study of psychological and neurophysiological processes in aging, Ruch's three verbal learning procedures were used in the present study. The findings of other studies in this series are reported elsewhere (2, 10). The present paper will consider only the comparison of two groups, distinct in age, in their performance on these procedures. The purpose is twofold: (*a*) To repeat Ruch's study and to retest his hypothesis; and (*b*) To see whether further understanding can be gained about age-related changes in cognitive functioning generally, and as they may account for performance in these learning tasks.

METHOD

Subjects

Two distinct age groups were used in the main study. The *young* group consisted of 24 *Ss* between

[1] This study was supported by the U. S. Army under Contract No. DA-49-007-MD-261 through the Medical Research and Development Board, Office of the Surgeon General, Department of the Army, Washington, D. C. We are grateful to Mr. Ben Grossman, Director, and the staff of the Drexel Home for their complete cooperation in providing subjects and facilities. Dr. Albert Caron and Mrs. Ann Gunther were responsible for much of the data collection and Dr. Helen Heath for the statistical analyses.

the ages of 22 and 33 years of age, and included 12 resident physicians with a mean age of 28.2 years and 12 graduate nurses with a mean age of 26.3 years. The *old* group consisted of 36 residents of an institution for the aged, evenly divided by sex within three age groups: 12 *Ss* between 65–74; 12 between 75–84; and 12 above the age of 85. The *Ss* were included in the older sample only if they were in good general health, free from organic brain pathology, and had acceptable corrected vision.

However, the total young group was distinctly superior to the old in education and intellectual level. Since one might reasonably expect intelligence, in general, to be related to performance on these tasks and, in fact, Ruch (13) had found this to be true, particularly within his older groups, it was felt that a more direct and exact test of the differences in learning could be made using groups equated for general intelligence. Since vocabulary correlates well with general intelligence as measured by omnibus batteries and itself shows relatively little change with age, the Vocabulary test of the Wechsler-Bellevue Scale was used (1, 4, 8, 15). Two matched groups of 16 *Ss* each were drawn from the larger sample. The young group had a mean unweighted vocabulary score of 26.8 with a standard deviation of 4.6; the old group 28.6 and 4.8, respectively. The young group consisted of 5 men and 11 women, with a mean age of 26.8; the old group of 8 men and 8 women, with a mean age of 78.1 years.

The results of this experiment will be reported for these selected groups matched in intelligence. The same analyses have been carried out, however, for the total groups and the results are quite similar, although in the main the differences between the larger groups are more extreme.

The Learning Tasks

As part of a larger battery of procedures, each *S* was individually administered a modified version of the learning tasks described by Ruch (13). As in his experiment, *Ss* had to learn three separate lists: 1. *Word Associates*, consisting of familiar, logically related word pairs such as Man—*Boy*; 2. *Nonsense Equations*, e.g., E × Z = *G*; 3. *False Equations*, e.g., 3 × 5 = *6*. In the present study the task was changed somewhat from Ruch's original procedure. After preliminary tests, the length of list was reduced from 10 to 8 items and the number of trials from 15 to 6 in order to make the task easier and the session less fatiguing, particularly for the older *Ss*.

TABLE 1

LISTS USED IN THE THREE LEARNING TASKS

	Word Associates	Nonsense Equations	False Equations
	Man-Boy (example)	E × Z = G (example)	3 × 5 = 6 (example)
1.	Stem-Bud	R × S = Q	2 × 5 = 8
2.	House-Visit	H × Y = J	6 × 3 = 5
3.	Room-Light	B × D = M	5 × 1 = 7
4.	Soft-Chair	A × M = B	3 × 1 = 1
5.	Tree-Flag	S × Q = H	3 × 3 = 4
6.	Nest-Owl	K × U = T	5 × 5 = 11
7.	White-Pink	W × L = S	2 × 4 = 9
8.	Walk-Car	L × B = D	3 × 4 = 2

Each list consisted of 8 units (Table 1) selected from the 10 used originally. Before each list was given *S* was first shown a sample pair and the task explained. After an initial presentation of the series, *S* was given 6 trials in which he was to anticipate the right-hand member of the pair when the left was first shown alone. The items were projected by a 35 mm. strip film projector in a semidarkened room, each frame remaining in view for 4 seconds. The left-hand member of the first pair appeared alone for 4 seconds, both parts for 4 seconds, then the left-hand member of the second pair, and so on. Eight seconds were allowed between trials and about 2 minutes between lists. To prevent serial learning, the order of items was changed on each trial.

The instructions used were substantially the same as Ruch's. The *S* was encouraged to guess if uncertain. In presenting the lists of Nonsense and False Equations the items were divided at the equal sign, e.g., first 3 × 5 = was shown, then 3 × 5 = *6*.

RESULTS

In Table 2 are presented the mean scores for each of the three tests. Since there were 8 items and 6 trials, the highest possible score for an individual is 48. Young *Ss* clearly perform better on each task. The difference between the groups is significant in each case at the $p < .001$ level on *U* test. Moreover, it is equally clear that for each group the Word-Associate learning was decidedly easier than either the learning of the Nonsense or False Equations, although, surprisingly, there is virtually no difference between the latter two.

Following Ruch's (13) analysis for determining variation in the amounts of deficit among the three types of material, we can compare the difference between old and young on each of the tests. For the Word Associates this difference in score is 13.81, for the Nonsense Equations 17.57, and for the False Equations it is 17.62 (Table 2). These values do not differ significantly from each other.

However, the difference between the groups in the learning of the three types of material

TABLE 2

MEAN LEARNING SCORES FOR YOUNG AND OLD GROUPS

Group	*N*	Mean Age	Word Associates (WA)	Nonsense Equations (NE)	False Equations (FE)	WA-NE / WA	WA-FE / WA
Young	16	26.8	44.19	25.00	24.81	43.59	44.21
Old	16	78.1	30.38	7.43	7.19	74.86	74.96
$Diff_{M_y - M_o}$			13.81	17.57	17.62		

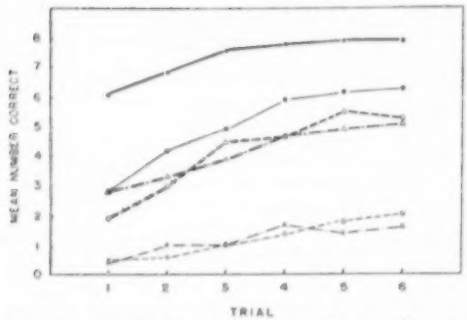


FIG. 1. LEARNING CURVES FOR THE OLD AND YOUNG GROUPS FOR EACH OF THE THREE TESTS

The curves for the young are heavily drawn, for the old lightly drawn. The three tests are distinguished: ——— Word Associates; - - - - Nonsense Equations; -·-·-·- False Equations.

can be viewed in another way. If we consider the proportional change between the tests it does appear that the older group has relatively poorer learning on both the Nonsense and False Equations when these are compared with the Word Associates. Subtracting the Word-Associate score and the Nonsense Equation score and taking this difference as a fraction of the Word-Associate score, the mean ratio $\left(\frac{WA - NE}{WA}\right)$ is 43.59 for the young *S*s and for the old 74.86. The difference between these values is significant, on *U* test, at the $p < .001$ level. Similarly, the ratio of Word-Associate score minus False Equation divided by Word Associate $\left(\frac{WA - FE}{WA}\right)$ is 44.21 for the young group and 74.96 for the old (Table 2); on *U* test significant at the $p < .001$ level. From this analysis we may fairly conclude that not only do the older *S*s learn less in each of the three tasks, but that they are proportionately more deficient in the learning of materials in which the facilitative effects of prior experience are minimized. However, we have not found evidence that interference materials are more difficult than nonsense materials.

The trial-by-trial learning curves for the two groups on each of the tests show the difference in performance already discussed to be visible throughout learning (Fig. 1). The curves for the Nonsense Equations and False Equations are virtually identical within each group and each of these is distinct from Word-Associates levels. It does seem, however, that the older group's curves for the Nonsense and False Equation tasks approximate linear functions while their values on the Word Assoicates more nearly describe the more usual negatively accelerated learning function. Particularly in the early trials there is proportionately less rise in score on the meaningless and interference materials. Using Edwards' (5) suggested analysis of variance for trend, the curves for both of these tasks differ significantly from that for the Word Associates ($p < .001$ in each case). This difference in the forms of the learning curves is suggestive of Lewis and Shepard's (11) finding of a negatively accelerated function for perceptual motor learning consistent with past experience and of a linear or positively accelerated curve in a task which did not utilize old habits.

Not only do the old and young groups differ in the amount of material learned, but they differ as well in the type of error made during the learning process. In Table 3 the total number of errors made by each of the groups on each of the three tasks is broken down into various types. In both the Word Associates and the Nonsense Equations there are three distinct types of error possible: *S* may not respond at all to the stimulus item (*No response*); he may give an associate which appears nowhere in the series (*Random response*); or he may respond with an item which although incorrect for the present stimulus item is appropriate elsewhere in the list (*Correct elsewhere*). For the False Equations list *S* can also give a response which represents a correct multiplication, or conceivably, he can add or subtract the two numbers in the stimulus equation. Although the completion of the multiplication is of particular theoretical interest since it may repre-

TABLE 3

PERCENTAGES OF VARIOUS TYPES OF ERRORS MADE IN WORD ASSOCIATE (WA), NONSENSE EQUATION (NE) AND FALSE EQUATION (FE) TESTS

Group	*N*	Mean Age	Test	Total No. of Errors	No Response	Random Response	Correct Elsewhere	Multiplication*	Addition*	Subtraction*
Young	16	26.8	WA	61	62	7	31			
			NE	368	45	7	48			
			FE	370	30	9	45	2	8	6
Old	16	78.1	WA	273	83	1	16			
			NE	649	92	1	7			
			FE	653	75	3	18	1	1	2

* These types of errors are only possible on the False Equation task.

sent the intrusion of the competing, albeit actually correct, response, addition or subtraction may also reflect habitual response to the implied structure of an arithmetical problem. Scoring these errors does, of course, involve interpretation since such response might as well be a *Random response* or, in a particular case, *Correct elsewhere*.

There is considerable difference in the pattern of errors between the two groups (Table 3). On all three tasks the typical behavior of the older person is not to respond, distinctly less often to give a response that is *Correct elsewhere*. Compared to the old group, young *Ss* have proportionally more of their errors of the latter type. They are also more prone to give *Random responses* and in the case of the False Equations to multiply, add or subtract the items, although there are few of these errors in either group.

Closer inspection of the trial-by-trial changes in the pattern of error gives further evidence of the difference between old and young learners. On all tasks on each successive trial the young group has relatively fewer of its errors in the *No response* class and an increasing proportion of *Correct elsewhere* errors. On the Nonsense Equations task, for example, of the total errors made by the young *Ss* on the first trial 57% are *No response* and 30% *Correct elsewhere*, which changes gradually to 30% *No response* and 66% *Correct elsewhere* by the sixth trial. However, the old group preserves roughly the same proportions on each trial. Thus, for the same task, these *Ss* start with 95% *No response* and 3% *Correct elsewhere* and by the last trial still have 92% and 8% of their errors in each of these classes. Although there is somewhat more tendency for the *No response* proportion to go down on the Word Associate and False Equation tasks, the pattern is basically the same. Thus, it is clear that simply not responding to a stimulus item is characteristic of the learning behavior of the older *Ss* throughout; it is less common in the younger *S* at any point and declines greatly as learning proceeds.

Lastly, how related are the performances of individuals on the three learning tests? The correlations among the total scores for the three tests within each of the age groups are given in Table 4. The results are surprising. For the young *Ss* there is a positive relationship among all the procedures, lowest between the Word Associates and Nonsense Equations, and quite high between the two forms of equations. However, the older group has its highest correlation between the Word Associates and Nonsense Equations and its lowest, and in a negative direction, between the Nonsense and False Equations. Thus, the patterns are opposite in the two groups. Moreover, these findings differ sharply from those of Ruch who found that all of the interest correlations were higher in his oldest group and highest among these was the correlation between the Nonsense and False Equations. Indeed, the relationship between these two tests was generally stronger than between any other pair in his experiments: +.404 in the young group; +.473 in the middle-aged; and +.833 in the old. We are at a loss to explain the differences in findings both between our two groups and between this experiment and Ruch's.

TABLE 4
CORRELATIONS AMONG LEARNING TASKS

	Young	Old
Word Associates vs. Nonsense Equations	.372	.506*
Word Associates vs. False Equations	.607*	.317
Nonsense Equations vs. False Equations	.740**	−.240

* Significant at $p < .05$.
** Significant at $p < .001$.

DISCUSSION

There is no doubt that older persons in this as in Ruch's (13) original study are less well able to master each of the learning tasks. As such, this finding is in agreement with a considerable body of material which shows a generally decreased ability to learn with advancing age (6, 12, 14). However, unlike Ruch, we have found virtually no difference between scores on the nonsense and interference materials and no evidence of a greater degree of deficit in the latter than in the former. As was noted earlier, the total amount learned by Ruch's *Ss* of all ages was greatest on the Word Associate task, second most on the False Equations and least on the Nonsense Equations, while deficit, i.e., difference in performance of young and old, was least for the familiar material, greater for the nonsense material, and greatest for the interference material. However, it may be that the differences, though seem-

ingly small, between Ruch's and the present procedure are sufficient to account for the obtained differences in results. Perhaps with longer lists and more trials the early findings would have been repeated, but in any case the differences in method do limit direct comparison of results.

What we have found is that, compared to the level of attainment on Word-Associate learning, the older persons show proportionately poorer learning on *both* the new and interference materials. In this sense, we can speak of greater deficit on both of these materials, although there is no difference between them. While it may be that older persons are uniquely prone to disturbed performance when required to reorganize existing habits, as has been suggested by a number of papers (7, 9, 14, 16), the evidence of the present experiment suggests that the older individual will perform less well in any situation in which he must deal with material which is in any way novel. When the task at hand can not be integrated with earlier experience, either because there is no precedent or existing precedent is inadequate or contradictory, learning suffers. The lessened capacity of the aged to deal with novelty was also found in heightened decision times at the beginning of a series when *S* had to judge a number of ambiguous stimuli (10).

Older and younger *S*s differ also in their behavior during learning. The younger, as we have noted, are more likely to make some response to the stimulus item, to offer hypotheses, whether correct or incorrect, and as learning proceeds to have an increasing proportion of their errors responses which, although correct elsewhere in the list, are inappropriately given to the particular stimulus. The older *S* will either respond correctly or not at all, and this behavior is consistent from trial to trial. In a study of the perception of closure the older *S*s also showed a greater tendency not to respond to Gestalt Completion Test items (2).

We believe that three factors jointly determine this behavior of the older *S*s, although from available evidence we cannot judge their relative contributions.

1. The older person is more cautious. Despite the injunction to guess, it may be that older persons require greater certainty before they are willing to report. Although *S* knows that the item is equally incorrect whether he gives no response or a wrong response, he apparently prefers the error of omission to that of commission. This tendency to inhibit response in the uncertain situation may reflect a more profound personality defense of the aged through which the recognition of inadequacy is avoided.

2. The older person requires more time for the integration of a response. A condition of the experiment which may account for some of the differences may be the rapid pacing of the items. As the experiment was conducted *S* had to read the stimulus item, select and report a response item within 4 seconds. Considerable work has shown that older persons require more time both for simple perceptual discrimination and in more complex decision-making (3). Moreover, requiring performance within arbitrarily short times can result in a significant reduction in the adequacy of that performance (16). Such findings would suggest that the processes of perceiving-judging-recalling-responding require more time for the older person than is available and that, perhaps, if the pacing were slower his behavior might differ less than that of younger individuals.

3. For the older person learning may principally involve the association of *discrete* stimulus-response combinations, and unlike the younger person, may not include a prior stage in which the response items are first learned as belonging *somewhere* in the series, before they are attached to particular stimuli. That the pattern of errors is generally the same on all three learning tasks and that there are no important changes in pattern from trial to trial further suggest that the acquisition of separate and distinct units may be a general characteristic of learning in the aged. Implicit in this view is the assumption that there are in general two patterns of learning: one in which there is first a more general learning from which particular associations are increasingly differentiated; and a second in which only the specific stimulus-response associations are formed. We might speculate that the second pattern, seen here in the behavior of aged persons, is generally characteristic of more constricted or rigid cognitive organization and might be found as well in the behavior of anxious or brain-damaged individuals.

From the first of these three views one assumes that the older learner may have available, but is unwilling to report, incorrect hypotheses. According to the second, the

formulation of such hypotheses can not be completed and reported within the time available. But on the last assumption neither caution nor time blocks response, rather there are fewer incorrect hypotheses available. The relative importance of these factors could be explored in further experiments in which, for example, *S*s might be required to make some response to each item and/or the time of presentation increased.

All of these factors, however, may be conceived as parts of the total aging process. In the aged there is a decreased ability to comprehend, organize, and integrate new experience both in perception and in memory. Integration is likely to be incomplete, often rigid or confused, and the integrative process requires greater time. Performance is most adequate where the individual can draw most heavily from past experience, most disturbed when novel behavior is involved. As a consequence the aged person seeks the security of the familiar, risks fewer failures by making fewer attempts, and so defends against the recognition of inadequacy.

SUMMARY

An old and a young group of *S*s, mean ages 78.1 and 26.8 years, were given three paired-associate learning tasks which differed in the degree to which prior experience might be expected to facilitate or block present learning. They consisted of (*a*) familiar word pairs, (*b*) nonsense equations, and (*c*) false equations.

Both groups performed best on the word-associate task, but there was little difference between the learning of nonsense and false equations within either group. On all three procedures the old group was significantly poorer, but they were proportionately more deficient in the learning of materials in which the facilitative effects of prior experience are minimized, i.e., the two forms of equations. However, they had no greater difficulty with the interference than with the nonsense material. The older person thus seems to be less capable in dealing with novel material which can not be readily intergrated with earlier experience.

Older and younger *S*s were also found to differ in their behavior during learning. Younger *S*s are more likely to offer hypotheses, whether correct or incorrect, and through learning to have an increasing proportion of their errors responses which though correct elsewhere in the series are incorrect for the particular stimulus. By contrast, the old *S*s respond correctly or not at all, and the proportion of "no response" errors remains high through learning. Three factors are discussed as joint determiners of this behavior: older persons are more cautious; they require more time to integrate response; and their learning may consist mainly of discrete stimulus-response associations, not preceded by more general learning. These factors are conceived as related parts of the total aging process.

REFERENCES

1. BABCOCK, HARRIET, & LEVY, L. *Test and manual of directions, the revised examination for the measurement of the efficiency of mental functioning*. Chicago: C. H. Stoelting, 1950.
2. BASOWITZ, H., & KORCHIN, S. J. Age differences in the perception of closure. *J. abnorm. soc. Psychol.*, 1956, **54**, 93–97.
3. BIRREN, J. E. Age changes in speed of simple responses and perception and their significance for complex behavior. In *Old age and the modern world*. Report of the Third Congress of the International Association of Gerontology. London: Livingstone, 1955. Pp. 235–247.
4. BOTWINICK, J., & BIRREN, J. E. Differential decline in the Wechsler-Bellevue subtests in the senile psychoses. *J. Geront.*, 1951, **6**, 365–368.
5. EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.
6. GILBERT, JEANNE G. *Understanding old age*. New York: Ronald, 1952.
7. HALSTEAD, W. A psychometric study of senility. *J. ment. Sci.*, 1943, **89**, 363–373.
8. HOWELL, R. J. Changes in Wechsler subtest scores with age. *J. consult. Psychol.*, 1955, **19**, 47–50.
9. KAY, H. Learning of a serial task by different age groups. *Quart. J. exp. Psychol.*, 1951, **3**, 166–183.
10. KORCHIN, S. J., & BASOWITZ, H. The judgment of ambiguous stimuli as an index of cognitive functioning in aging. *J. Pers.*, 1956, **25**, 81–95.
11. LEWIS, D., & SHEPARD, A. H. Prior learning as a factor in shaping performance curves. *Proc. nat. Acad. Sci.*, 1951, **37**, 124–131.
12. MCGEOCH, J. A., & IRION, A. L. *The psychology of human learning*. New York: Longmans, Green, 1952.
13. RUCH, F. L. The differentiative effects of age upon human learning. *J. gen. Psychol.*, 1934, **11**, 261–268.
14. THORNDIKE, E. L., BREGMAN, ELSIE O., TILTON, J. W., & WOODYARD, ELLA. *Adult learning*. New York: Macmillan, 1928.
15. WECHSLER, D. *The measurement of adult intelligence*. (3rd Ed.) Baltimore: Williams & Wilkins, 1944.
16. WELFORD, A. T. *Skill and age*. London: Oxford Univer. Press, 1951.

Received March 1, 1956

PERSONALITY, PROBLEM SOLVING, AND THE EINSTELLUNG EFFECT[1]

BRENDAN A. MAHER

Northwestern University

ADORNO (1) and others have reported a pattern of personality characterized by certain cognitive and behavioral tendencies among which is included rigidity of set and outlook (1, p. 561). The concept of rigidity has been employed in many ways, a comprehensive summary of which is to be found in a paper by Fisher (6), but a review of its uses is unnecessary here. This study is concerned with rigidity as defined by Adorno and his colleagues, ("pre-conceived categorizing, inaccessible to new experience") in relation to certain other aspects of personality, in particular pseudo-scientific or antiscientific tendencies in interpreting individual behavior.

The "ethnocentric" individual is described by the same writers as exhibiting a cognitive personality organization that includes "implicit denial of personality dynamics, ready explanation by accidental factors, heredity, etc., superstition." This method of organizing perceptions of the self and others is defined as antiscientific and is contrasted with the tendency to explain human behavior in dynamic sociopsychological terms.

Rokeach (12) has used the familiar water-jar problems to demonstrate that rigidity of set in problem solving is found more frequently in "high" ethnocentric groups than in "low" groups. His findings have been challenged by Luchins (11) and Brown (2); the later concluding that this relationship existed only "... in ethnocentric *S*s when they felt an anxious concern over achievement."

Against this background the present study was designed to test two simple hypotheses: (*a*) *S*s who reveal antiscientific or nondynamic tendencies in their perceptions of human behavior are more subject to the Einstellung effect in cognitive problem solving than *S*s who use dynamic psychological concepts in interpreting human behavior. (*b*) This relationship is more evident when the *S*s are placed in a situation likely to give rise to achievement anxiety.

The use of "antiscientific" bases for evaluating individual behavior has been described in an unpublished study by Sullivan (13). He used a method in which *S*s were required to sort and group photographs of people on any basis that they might wish to use. The sorting categories reported by the *S*s were then classified as objective or nondynamic concepts (e.g. hair-color, physical characteristics in general, etc.), or as inferential (dynamic) concepts, the latter implying some attempt to use inferred psychological or behavioral characteristics. His findings were the reverse of those to be predicted on the basis of the previous studies in ethnocentrism. It was suggested post facto that in this experiment the "scientific" or objective approach would properly utilize nondynamic categories, as the judgment of personality from photographs is essentially an unscientific procedure based on a willingness to infer personality attributes on the basis of pictured, external characteristics.

To avoid this ambiguity of interpretation, the present study used a measure derived from Kelly's Role Construct Repertory Test (RCRT) (9). This procedure presents *S* with the task of evaluating particular individuals with whom he is personally acquainted, and may therefore be assumed to elicit the perceptual categories that he uses in everyday life. In the test, *S* lists a number of persons of his acquaintance, on the basis of definitions (Kelly's "role-titles") provided for him. He is then required to consider particular trios of these individuals in turn. Each trio is divided by *S* into a single and a pair of individuals on the basis of some perceived similarity between the two members of the pair which is not shared by the third (single) member. For each trio, *S* reports the basis upon which the division was made. The list of definitions is the same for each *S*, but the persons actually listed will, of course, be different. For any particular purpose the number and nature of the definitions may

[1] The data upon which this paper is based are taken from a dissertation submitted by the writer to the Graduate School of the Ohio State University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. Particular acknowledgments are due to Dr. George Kelly for his advice and encouragement

be altered, and the number of trios may be of any size within the limits of combinations permitted by the number of definitions.

One estimate of the reliability of the RCRT has been derived by Hunt (8). He administered the test to a sample of college students followed one week later by a similar administration but using different role titles. A mean percentage agreement of 70.1 was obtained. As responses were only regarded as in agreement when they were identical, it is apparent that any broader definition of similarity could not reduce Hunt's percentage and would be likely to increase it.

METHOD

Subjects and Procedure

Undergraduates in an elementary mental hygiene class were divided into two equal groups.

Achievement anxiety group. These were students in one section of the course. At the beginning of the course they completed the RCRT. Five weeks later a mid-quarter examination was held in which all *S*s participated as a course requirement. At the next class meeting the group was informed that the examination results had been very unsatisfactory and that as an aid to grading them two simple tests of "mental alertness" would be administered immediately. The problem-solving tests were then administered in group fashion. At the end of the class hour their graded examination papers were returned to the group and the experiment explained. This procedure was intended to demonstrate that the problem-solving tests had not actually affected their examination grade and to eliminate any residual anxieties. It was *E*'s impression that during the test session the motivation of the group was sufficient to justify the inference that achievement anxiety was a relevant variable.

Control group. This group, which comprised another section of the same course, completed the RCRT as above. The problem-solving tasks were administered in group fashion during a regular class meeting. They were presented with standard instructions as part of the demonstrations used regularly in the course.

This procedure was repeated over two academic quarters because of the attrition rate associated with errors in the completion of the printed problems. Complete data were obtained from 62 *S*s, 31 in each group.

Instruments

Verbal maze problems. This series of problems was identical in content and sequence with that reported by Cowen[1] (5). Each maze was ditto-printed on a separate page. The pages were stapled together in such a fashion that it was necessary to solve each maze in turn and then detach it before being able to inspect the next problem in the series. Following Cowen's method, the problems were presented in sequence commencing with two practice problems; one control problem (which could be solved by a long or a short method of solution); five Einstellung problems soluble only by the long method; four crucial problems which, like the control problem, were amenable to either method of solution; one extinction problem, not soluble by the long method; and finally one more crucial problem. The Einstellung score is the number of crucial problems solved by the long method, and could attain a maximum value of 5.

[1] The complete set of mazes was supplied to the writer by Cowen in a personal communication.

TABLE 1
EINSTELLUNG PROBLEMS: WATER-JARS SERIES

Problem Number	Sizes of Jars Available			Quantity to be Obtained	Type of Problem
A	39	4		31	2 jar practice
B	30	40	3	4	3 jar practice
1	27	74	20	7	Control
2	42	71	5	19	1st Einstellung
3	50	81	7	17	2nd Einstellung
4	23	57	9	16	3rd Einstellung
5	27	71	14	16	4th Einstellung
6	35	77	17	8	5th Einstellung
7	21	51	9	12	1st Crucial
8	32	79	15	17	2nd Crucial
9	37	90	16	21	3rd Crucial
10	31	75	13	18	4th Crucial
11	19	65	7	26	Extinction
12	28	65	9	19	5th Crucial

Water-jar problems. These problems were presented in the same control-crucial sequence as the verbal mazes. The problems used are given in Table 1 below. Administration of the problems was in booklet form to the groups of *S*s in a manner identical with that used in the verbal mazes. To take into account the possibility that perception of the correct solution in one series of problems may transfer to the solving of the second series, one half of each group commenced with the verbal mazes and then went on to the water-jars problems. The sequence was reversed for the other half of each group. Thus while transfer effect was not eliminated, it is presumed to have influenced the group mean scores equally for both problems.

Role Construct Repertory Test. The RCRT was administered with 11 definitions and 22 triads. The definitions used were as follows:

1. "Myself"
2. Father or surrogate
3. Mother
4. Brother
5. Sister
6. Opposite sex peer most understood
7. Opposite sex peer least understood
8. College professor most understood
9. College professor least understood
10. Same sex peer most understood
11. Same sex peer least understood

The triads were selected randomly with the restriction that each defined figure was to be categorized six times.

From each *S*, a total of 22 category responses were obtained, with the possibility that any category might appear more than once. The categories used by each *S* were typed on separate cards and numbered to identify the protocol from which they were taken. As a

result of the considerable number of duplications found within and between protocols, the total of categories was well below the possible maximum of 1,364. The cards were then inspected and sorted into two groups by each of 3 judges in turn, on the basis of the following criteria.

Nondynamic categories. A group including all those categories based upon: (*a*) Physical appearance, health, height, weight, etc.; (*b*) Kindred relationship, degree of acquaintance, etc.; (*c*) Legal status, sex, age, occupation; (*d*) Geographical, cultural, or social group membership; (*e*) Presence or absence of a specific skill or achievement.

Dynamic categories. A group including all categories referring to personality in terms of needs, motives, drives, characteristic behavior patterns, etc. Examples would be "aggressive," "intelligent," "ambitious," "anxious" and the like.

The judging was on a forced-choice basis, as each category card was to be placed in one of the two groups. Where differences arose between the judges' evaluations, the majority decision was accepted. Each judge operated independently, the collating of judgments being done by the experimenter. Interjudge agreement was computed in terms of percentage of agreement on the total cards judged. The lowest agreement between any two judges recorded was 91.1; the highest was 99.0. Under the circumstances it seems reasonable to conclude that the criteria are reliable as bases for judgment.

The judging completed, each *S* was then scored with respect to the number of times which he used nondynamic categories to discriminate between figures. The maximum possible source was 22. Scores obtained ranged from 0 to 19. A range of 0 to 3 included the 31 lowest scores, while the high scores ranged from 5–19. The mean scores of these two groups were 1.8 and 9.3 respectively.

RESULTS

The mean Einstellung scores of both groups on the two problem-series are presented in Table 2. Analysis of these data revealed a statistically significant difference in respect to the achievement-anxiety variable on the water-jars problems. There were no other significant differences between means. In no instance did an *S* use the long method for solution of the final (postextinction) crucial problem.

TABLE 2

MEAN EINSTELLUNG SCORES IN RELATION TO ACHIEVEMENT ANXIETY AND TO USE OF NONDYNAMIC (ND) CATEGORIES

Groups	*N*	Einstellung Test Means	
		Water jars*	Mazes**
Achievement anxiety			
High ND	17	2.13	2.20
Low ND	14	2.25	2.07
Control			
High ND	14	1.55	1.80
Low ND	17	1.44	1.63

* For achievement anxiety vs. control, $t = 2.29$, $p < .05$; for High ND, $t = 1.99$ (n.s.)

** For achievement anxiety vs. control, $t = 1.97$ (n.s.); for High ND vs. Low ND, $t = 1.3$ (n.s.)

In view of the fact that the High Nondynamic (ND) grouping includes a wide range of scores, it was deemed advisable to ensure that the differences in problem scores were not being contaminated by initial differences in ND scores between the Control High ND and the Achievement-Anxiety High ND subgroups. Therefore the 20 highest and 20 lowest ND scores were selected and their distribution between the four groups was tabulated. A chi square derived from this distribution was nonsignificant. It is therefore assumed that the differences found between groups can be attributed to the obvious variables and not to chance maldistribution of the High ND *S*s between the control and experimental conditions.

DISCUSSION

The data obtained in this study do not support the hypothesis that individuals who show nondynamic tendencies in their perceptions of other people are more susceptible to the Einstellung effect in cognitive problem-solving situations. Some support is offered for the conclusion that the presence of achievement anxiety creates this susceptibility in *S*s irrespective of their tendencies to dynamic or nondynamic perception in interpersonal relations. No evidence of any interaction effects was found and the obtained difference was significant only for the water-jars problems.

These findings appear to be relevant to the broader issue of generality in rigidity. The issue is, essentially, whether rigidity is to be regarded as a dimension of a given person's behavior that underlies many diverse actions, from social relationships to mathematical calculations, or as relatively specific to a given situation. The latter alternative implies that rigid behavior might be elicited from anyone in that situation. It appears that the data reported above support the latter rather than the former position. In this respect they run counter to the findings of Brown in a similar experimental situation (2). Brown reports a significant correlation between scores on the F scale and Einstellung scores when *S*s were

working under conditions designed to arouse achievement anxiety, but not under "relaxed" conditions, with the implication that rigidity functions as a "preferred" defense for some persons, being elicited by any stimulus perceived as threatening to status or prestige. In the absence of such threat, the rigidity would be either minimal or absent.

The fact that this relationship was not found in the present study may be due to a number of factors. First, Brown did not use a control problem in his water-jars series as he found that its omission maximized the Einstellung effect. The fact that a control problem was used in the present study may contribute to an explanation of the different results. Whether or not to use a control problem depends upon the necessity of observing the susceptibility of *Ss* to set induced *within* the confines of the laboratory situation. A second factor is related to the arousal of achievement anxiety. In Brown's study, "the experimenter (*E*) maintained an extremely grave, aloof manner, dressed formally and conservatively and repeatedly cautioned *Ss* against looking at their test materials ahead of time," and suggested that "the tests to be taken were measures of intelligence and motivation and the results were of great importance." The present experiment, in contrast, related performance on the tests to a rather specific punishment-reward situation—namely a grade on a course in which *Ss* were enrolled, and may have created a higher level of achievement anxiety. Perhaps when the anxiety is sufficiently intense, even the most flexible of individuals may show rigid behaviors. When this happens, differences between individuals which may be evident in less threatening situations become obscured. Brown comments that "when the authoritarian is presented with the arithmetic problems in a very relaxed social situation he solves them as easily as the equalitarian." To this might be added the comment that when the equalitarian is presented with the arithmetic problems in a situation that is *sufficiently* anxiety arousing, he solves them no more easily than the authoritarian. The relationship between stress, anxiety, and rigidity has already been reported in the literature (3, 4, 7, 10).

The problem of generality in rigidity may be considered in terms of the dimensions along which rigidity is assumed to generalize. If situations involving two or more pencil-and-paper cognitive problems show significant correlations within a given group of *Ss* then, it is possible to infer a factor of rigidity operating in *Ss*' responses to this type of problem. It may be possible to predict that rigidity will be elicited in a given *S*'s response to other similar paper-and-pencil cognitive problems. Such evidence seems to be an inadequate basis, however, for predicting behavior in *social* situations. This position resembles that of Cowen, Wiener, and Hess (5) who obtained significant correlations between the water-jars problems and alphabet-mazes, but point out that such correlations cannot be regarded as evidence that there is a "relationship between rigidity and personality" except in the precise experimental situation in which the correlations were obtained. With this point of view the present writer finds himself in close agreement.

SUMMARY

Two groups of *Ss*, 31 in each group, completed two Einstellung-test problem series. One group was tested under standard conditions, the other under conditions calculated to arouse achievement anxiety. The *Ss* were then scored on a variable of dynamic or nondynamic perceptions of other people as defined by Adorno and his co-workers (1). The measure was derived from a modification of Kelly's Role Construct Repertory Test. This variable was found to be unrelated to the Einstellung-test scores, whereas a significant relationship was found to achievement anxiety. The results are interpreted as generally favorable to the view that situational or field variables play a significant role in the production of rigid behavior.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. BROWN, R. W. A determinant of the relationship between rigidity and authoritarianism. *J. abnorm. soc. Psychol.* 1953, **48**, 469–476.
3. CHRISTIE, J. R. The effects of frustration on rigidity in problem solution. Unpublished doctor's dissertation, Univer. of California, 1949.
4. COWEN, E. L. The influence of varying degrees of psychological stress on problem-solving rigidity. *J. abnorm. soc. Psychol.*, 1952, **47**, 512–519.
5. COWEN, E. L., WIENER, M. & HESS, JUDITH. Generalization of problem-solving rigidity. *J. consult. Psychol.*, 1953, **17**, 100–176.

6. FISHER, S. Patterns of personality rigidity and some of their determinants. *Psychol. Monogr.*, 1950, **64**, No. 1 (Whole No. 307).
7. HARRIS, R. A. The effects of stress on rigidity of mental sets in problem solution. Paper read at A.P.A., Washington, D. C., Sept., 1952.
8. HUNT, D. E. Studies in the role concept repertory: conceptual consistency. Unpublished master's thesis, Ohio State Univer., 1951.
9. KELLY, G. A. *The psychology of personal constructs.* New York: Norton, 1955.
10. KENDIG, ISABELLE, V. Studies in perseveration: II. Determining factors in the development of compulsive activity. *J. Psychol.*, 1937, **3**, 231–246.
11. LUCHINS, A. On recent usage of the Einstellung-effect as a measure of rigidity. *J. consult. Psychol.*, 1951, **15**, 89–94.
12. ROKEACH, M. Generalized mental rigidity as a factor in ethnocentrism. *J. abnorm. soc. Psychol.*, 1948, **43**, 259–278.
13. SULLIVAN, P. Conceptual properties governing the categorization of people. Unpublished doctor's dissertation, Univer. of California, 1950.

Received March 5, 1956.

ON THE RELATION BETWEEN THE PROBABILITY OF A WORD AS AN ASSOCIATION AND IN GENERAL LINGUISTIC USAGE

DAVIS HOWES

Center for Advanced Study in the Behavioral Sciences[1]

THE word-association experiment, developed primarily as a clinical tool, is now being used extensively as a method for the quantitative analysis of linguistic behavior (1, 3, 5, 6). Its value in such investigations depends to a considerable degree upon the extent to which results of the experiment can be generalized to language emitted under other, more natural conditions. The criticism is often advanced, for example, that the specialized nature of the instructional set and other conditions attendant upon the experimental situation invalidates the association experiment as a technique for studying general linguistic usage.

The question can be resolved empirically only by comparing specific properties of language produced in the association experiment with corresponding properties of language produced under conditions free of the constraints of the experimental situation. One property that has been studied is the distribution of word-association frequencies, which Skinner (7) has shown to approximate Zipf's law as closely as do the distributions of word frequencies taken from samples of continuous discourse (11). The comparison is here extended from the form of the distributions to the actual frequencies with which individual words occur in both types of sample.

To perform this comparison we need a listing of all different words with their frequencies of occurrence in general usage in one column and their over-all frequencies of occurrence as responses in the word-association experiment in a second parallel column. If the correlation between the frequencies in the two columns is perfect, the language of the word-association experiment may be said to be *stochastically equivalent* to the language of general usage. In that event we could conclude that the special conditions introduced by the association experiment do not affect a person's selection of words. The effect of a stimulus word in the association experiment, on this hypothesis, is merely to redistribute the factors underlying the emission of words without changing them. If, on the other hand, the frequencies in the two parallel columns are so disparate that it is impossible to predict one set of frequencies from the other, we would be forced to conclude that the experimental situation introduces fundamental changes into the processes underlying a person's selection of words. An empirical determination of the correlation between the two columns will permit us to decide which of these two conceptions of the association experiment applies.

[1] On leave from Operational Applications Laboratory, AF Cambridge Research Center. Present address: Department of Economics and Social Science, M.I.T.

DEFINITIONS

An explicit statement of the problem requires more careful definition of concepts than has been traditional in this field. In this section the concept of stochastic equivalence of language samples is developed from definitions based upon certain idealized linguistic experiments.

We restrict ourselves to the written language produced by some specified population of persons O. It is assumed that the written language emitted by O can be decomposed into sequences of occurrences of a set of discrete spatial patterns or *letters*. Any permutation of letters bounded by spaces is a *word*, and the set of all possible words for a given set of letters is the *vocabulary*. Thus for English the vocabulary is the denumerable set of words A, B, $\cdots$, Z, AA, AB, $\cdots$, ZZ, AAA, AAB, $\cdots$, ZZZ, $\cdots$.

A word is said to be emitted *in the free situation* if at the time of its emission the behavior of O is in no way affected by the experimenter. An occurrence of the jth word of the vocabulary is written w_j if it is emitted in the free situation.

Consider a fixed time interval T, called the *base interval*, during which a random sample from the population O emits words in the free situation. The probability of obtaining an occurrence of the jth word in a random draw from the sample of words emitted in T is the *base probability* of that word, $P(w_j)$. The base

probabilities of all words in the vocabulary define a vector $\mathbf{P}$,

$$\mathbf{P} = [P(w_1),\ P(w_2),\ \cdots,\ P(w_j),\ \cdots], \qquad \Sigma_j P(w_j) = 1, \tag{1}$$

which is called the *base probability vector*. This vector is fixed uniquely by a population O and a base interval T.

Similar definitions apply to the written language produced in a word-association experiment. Let another random sample from the same population O be given a standard association experiment (4, p. 38) during randomly-selected segments of the interval T. The symbol r_j will be used to represent an occurrence of the jth word as a response (association) in such an experiment in order to distinguish it from a word emitted in the free situation. A stimulus presentation of the ith word by the experimenter is denoted by s_i. The probability of obtaining an occurrence of the jth word in a random draw from the sample of associations given in response to a stimulus presentation of the ith word is the *associative probability* of the jth following the ith word, $p_{s_i}(r_j)$. The associative probabilities of all words in the vocabulary following presentation of the ith word define a vector $\mathbf{p}_i$, called the *associative probability vector* for the ith word,

$$\mathbf{p}_i = [p_{s_i}(r_1),\ p_{s_i}(r_2),\ \cdots,\ p_{s_i}(r_j),\ \cdots], \qquad p_{s_i}(r_i) = 0, \qquad \Sigma_j p_{s_i}(r_j) = 1. \tag{2}$$

The ith term of this vector, $p_{s_i}(r_i)$, is always zero, since the standard experimental procedure does not permit the subject to give the stimulus word as his response.

To establish the connection between the associative vector $\mathbf{p}_i$ and the base vector $\mathbf{P}$ we consider the vector of conditional probabilities of words emitted in the free situation,

$$\mathbf{P}_i = [P_{w_i}(w_1),\ P_{w_i}(w_2),\ \cdots,\ P_{w_i}(w_j),\ \cdots], \qquad \Sigma_j P_{w_i}(w_j) = 1. \tag{3}$$

Since

$$P(w_j) = \Sigma_i P(w_i) P_{w_i}(w_j), \tag{4}$$

the base vector $\mathbf{P}$ represents the weighted average of the vectors $\mathbf{P}_i$,

$$\mathbf{P} = \bar{\mathbf{P}}_i = \Sigma_i P(w_i) \mathbf{P}_i. \tag{5}$$

In order to obtain a comparable quantity for words emitted in the association experiment we define a *base associative probability vector* as the identically weighted average of the vectors $\mathbf{p}_i$,

$$\mathbf{p} = \bar{\mathbf{p}}_i = \Sigma_i P(w_i) \mathbf{p}_i. \tag{6}$$

One can think of the base associative probabilities as the frequencies of occurrence of words as responses in an association experiment in which the stimulus words represent a random sample of the words the experimental subjects would have emitted during the same interval of time had they been left in the free situation. In practice, no word is presented as a stimulus more than once in a given association experiment. Each associative probability vector $\mathbf{p}_i$ therefore has to be estimated from responses made to a single experimental presentation of the ith word. It is therefore convenient to compute the base associative vector from the product of the matrix of associative probabilities $[\mathbf{p}_1,\ \mathbf{p}_2,\ \cdots,\ \mathbf{p}_i,\ \cdots]$ post-multiplied by the column vector $\mathbf{P}$

$$\mathbf{p} = \bar{\mathbf{p}}_i = [\mathbf{p}_1,\ \mathbf{p}_2,\ \cdots,\ \mathbf{p}_i,\ \cdots]\mathbf{P}. \tag{7}$$

The concepts introduced at the outset of the paper can now be given more precise definition. *Stochastic equivalence* between the language emitted in the association experiment and the language emitted in the free situation is defined by equality of the vectors $\mathbf{P}$ and $\mathbf{p}$. When these vectors are defined as in Equations 5 and 6 the equivalence is said to be of the *first order*. First-order stochastic equivalence thus implies that, for every pair (i, j) of words in the vocabulary,

$$\Sigma_i P(w_i) P_{w_i}(w_j) = \Sigma_i P(w_i) p_{s_i}(r_j). \tag{8}$$

Stochastic equivalence of any higher order can be defined by substituting conditional and associative probabilities of a higher order for those of the first order in Equations 2 and 3. Stochastic equivalence of the second order, for example, would imply in analogy to Equation 8 that

$$\Sigma_{ij} P(w_{ij}) P_{w_{ij}}(w_k) = \Sigma_{ij} P(w_{ij}) p_{s_{ij}}(r_k) \tag{9}$$

for all triples (i, j, k) of words in the vocabulary. In the present study we shall be concerned exclusively with first-order relationships. It is important, however, to recognize the possibility that the results obtained for the first order may not apply to higher-orders.

The equation $\mathbf{P} = \mathbf{p}$ is trivial if the associative probabilities of all word pairs are equal to their corresponding conditional probabilities, i.e., if for all pairs (i, j) of words,

$$p_{s_i}(r_j) = P_{w_i}(w_j),$$
$$p_{s_k}(r_k) = 0, \qquad \Sigma_j p_{s_i}(r_j) = 1, \qquad [10]$$
$$P_{w_k}(w_k) = 0, \qquad \Sigma_j P_{w_i}(w_j) = 1.$$

Now, there are no data from which the conditional probabilities $P_{w_i}(w_j)$ can be estimated accurately. But inspection of the data from any word-association experiment is alone sufficient to show that in fact associative probabilities differ greatly from conditional probabilities for at least some word pairs. To cite one example, the associative probability of the word *light* following the stimulus word *dark* is found experimentally to be about 0.4, while the probability that *light* follows *dark* in the free situation (i.e., the probability of the sequence *dark light* over the probability of *dark*) is clearly of a lower order of magnitude. Thus the real question at issue is whether the average associative probabilities are equal to the average conditional probabilities even though the individual associative and conditional probabilities differ greatly.

The method used here to test for stochastic equivalence is to calculate the coefficient of correlation between empirical estimates of $P(w_j)$ and $\Sigma_i P(w_i) p_{s_i}(r_j)$ for as large a sample of words as possible. A correlation coefficient of one implies perfect stochastic equivalence; a correlation of zero, perfect stochastic independence. The square of the correlation coefficient measures the proportion of the variance of the empirical data that is due to the relationship between the two sets of frequencies. It is thus a measure of the extent to which the hypothesis of stochastic equivalence accounts for those data. The square of the correlation also indicates the precision with which the probabilities of words in the free situation can be estimated from the laboratory situation of the association experiment. In this sense, it may be interpreted as an index of the potential value of the association experiment as a general laboratory method for the statistical analysis of language.

MEASUREMENTS

Measurements of base probabilities and associative probabilities for the same population of subjects have not been published. For base probabilities the most extensive set of data published is the Lorge magazine count (8). This is an analysis of 4.6 million words of text from five popular American magazines issued during the period 1927–1938. For associative probabilities the standard tables are still those published by Kent and Rosanoff in 1910 (4), which list the associations spoken to each of 100 stimulus words by a population of 1,000 subjects, including 500 hospital workers, 200 professional persons and 150 high school students. Since the two sets of data differ with respect to population of subjects, time when measured, and mode of response (written or spoken), an estimate of the correlation between base probability and base associative probability based on these tables cannot be expected to approach unity even in the case of stochastic equivalence.

The Kent-Rosanoff tables list each combination of letters as a separate word, but the published version of the magazine count combines the entries for certain inflected forms (e.g., plurals ending in *s*, verb forms ending in *ed* or *ing*) with the root forms. The corresponding entries in the Kent-Rosanoff tables have been combined in the present analysis to make the two sets of frequencies comparable.

The stimulus words used in the Kent-Rosanoff study constitute a very unrepresentative sample of the set of all possible words. All are nouns and adjectives, for example, and all are words of medium frequency (their median log frequency is 2.6 with a standard deviation of 0.5 log unit). Weighting each associative frequency by the magazine-count frequency of the stimulus word appears to be of limited value in such a case. On the other hand, the weighting procedure greatly increases computational labor. Computation of base associative probabilities has therefore been simplified by using the approximation

$$\Sigma_i P(w_i) p_{s_i}(r_j) \simeq 0.01 \Sigma_i p_{s_i}(r_j) \qquad [11]$$

in which the 100 stimulus words receive equal weights.

The notation may be simplified somewhat for the purpose of describing the data by dropping the indicators w, s, and r which distinguish the various functions of words in the experiments. The base probability is then denoted by $P(j)$ and the base associative probability by $\Sigma_i P(i) p_i(j)$. The empirical estimates

of these theoretical quantities will be referred to as the *magazine-count frequency* $F(j)$ and the *summed Kent-Rosanoff frequency* $\Sigma f_s(j)$, respectively.

ANALYSIS OF DATA

Figure 1 is a plot of the relation between magazine-count frequency and summed Kent-Rosanoff frequency on logarithmic coordinates. The numbers bordering the scatter plot represent the numbers of words having frequencies of zero, whose logarithms are therefore negative infinite. A significant positive correlation is evident from the figure, but the relationship is by no means rectilinear over the entire range. The form of the regression lines, as well as the magnitude of the correlation, is therefore of interest. Both types of analysis, in turn, require that the form of the conditional distribution functions be known.

Variation of base associative probability with base probability constant. Consider first the conditional distributions in the vertical dimension of Fig. 1. These distributions show what proportion of words having a given magazine-count frequency occur 0, 1, 2, ⋯ times as responses in the Kent-Rosanoff tables. Figure 2-A presents several such distributions on logarithmic-normal coordinates. A straight line on these coordinates indicates that the logarithms of the variable plotted on the abscissa are normally distributed with a mean indicated by the intercept of the line and a standard deviation inversely proportional to its slope. The sets of points in Fig. 2-A are well described by a set of lines that are both straight and parallel, indicating that the conditional distributions of log base associative probability for given base probability are both normal and homoscedastic. (Actually, there is a slight increase in the standard deviation of log $\Sigma f_s(i)$ at the highest magazine-count frequencies, but this is due to the use of wider class-intervals in that range in order to obtain distributions of more than 100 cases.)

The regression of summed Kent-Rosanoff frequency on magazine-count frequency is shown by the solid circles in Fig. 3-A. The coordinates are logarithmic. Medians have been substituted for means in the regression analysis because the distributions include zero frequencies the logarithms of which are infinite. The regression line has a maximum at a magazine-count frequency of about 800 and falls off sharply on both sides of that point.

It may be observed that the regression line in Fig. 3-A covers only about half the total range of log $F(j)$. Words that are rare in the magazine count tend not to occur at all in the Kent-Rosanoff tables, which comprise a much smaller sample. When the magazine-count frequency is small (<150), the median summed Kent-Rosanoff frequency is therefore zero and the median log $\Sigma f_s(j)$ is negative infinite. At the very highest magazine-count frequencies more than half the words again fail to occur in the Kent-Rosanoff tables, so that the median log $\Sigma f_s(j)$ is negative infinite in that range too.

An estimate of the regression line over the full range can be found by using a theorem which states that the line connecting a given fractile of each of a set of normal, homoscedastic distributions is parallel to a line connecting any other fractile of the same set of distributions.[2] Since the highest fractiles of the conditional distributions of log $\Sigma f_s(j)$ are finite for all magazine-count frequencies, the slope of the regression line can be estimated from the slope of a line connecting one of those fractiles. In Fig. 3-A the 95th and 75th percentiles of the log $\Sigma f_s(j)$ distributions are indicated by crosses and triangles, respectively. These data indicate a regression line that is rectilinear and positive for magazine-count frequencies of less than 800 (log $F(j) < 2.9$). (The leveling-off of the 95th percentiles for log $F(j) < 1$ is due to an artifact explained in the following paragraph.) Above that point the regression line undergoes a sudden reversal until at the highest magazine-count frequencies its slope approaches $-\infty$. The significance of this reversal will be taken up in a later section.

The leveling-off of the 95th percentiles for log $F(j) < 1$ is due to the omission of a number of words in that range. The magazine count is

[2] A proof may be indicated for a set of normal distribution functions $G_1(x)$, $G_2(x)$, ⋯, $G_n(x)$ having standard deviations $\sigma_1 = \sigma_2 = \sigma_n = \sigma$. Let a and b be two proportions, $0 \leq a < b \leq 1$, for which the fractiles of the various distribution functions are $x_1(a)$, $x_2(a)$, ⋯, $x_n(a)$ and $x_1(b)$, $x_2(b)$, ⋯, $x_n(b)$, respectively. If $h(a)$ and $h(b)$ are the corresponding fractiles of the standardized normal distribution function, we can write $x_1(b) - x_1(a) = [h(b) - h(a)]\sigma = x_2(b) - x_2(a) = x_n(b) - x_n(a)$. A line connecting the points $x_1(a)$, $x_2(a)$, ⋯, $x_n(a)$ will therefore lie $[h(b) - h(a)]\sigma$ units of x below a line connecting the points $x_1(b)$, $x_2(b)$, ⋯, $x_n(b)$. Hence the two lines will be parallel.

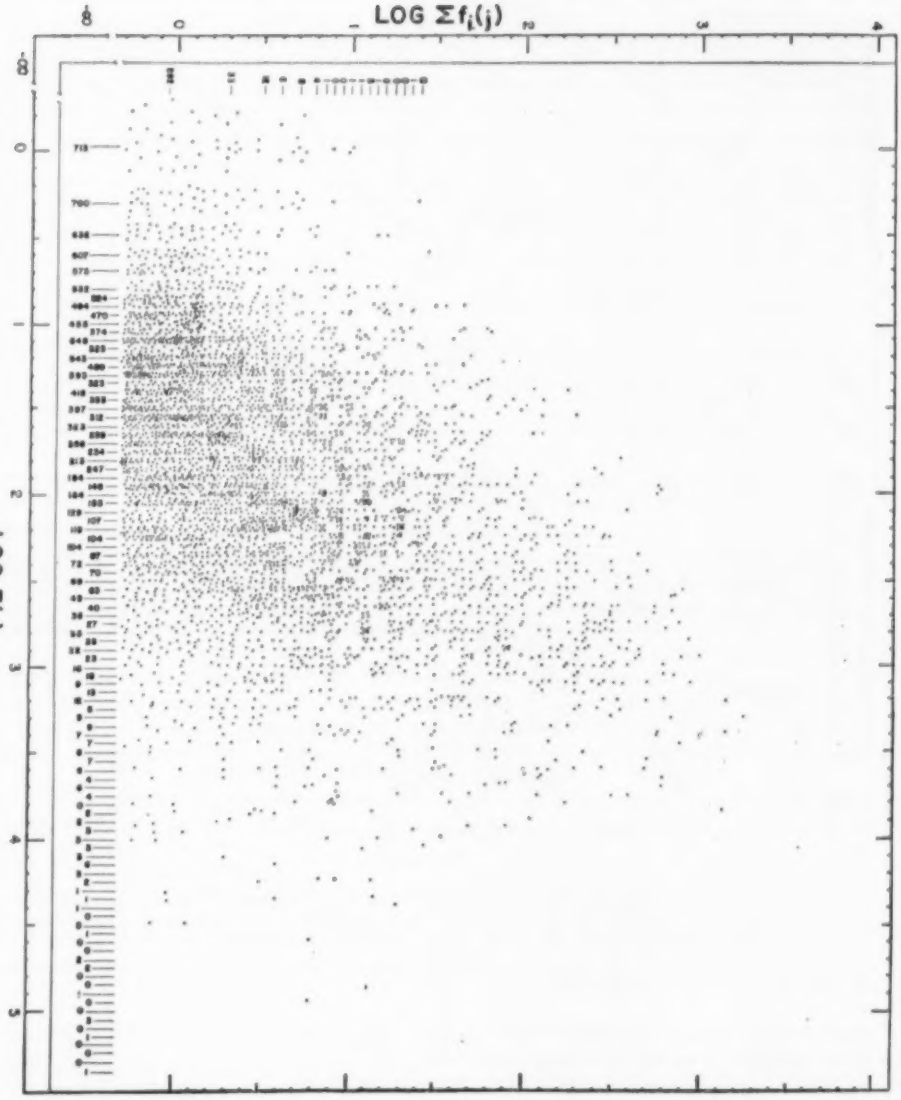


FIG. 1. SCATTERPLOT SHOWING THE RELATIONSHIP BETWEEN SUMMED KENT-ROSANOFF FREQUENCY, $\Sigma f_i(j)$, AND MAGAZINE-COUNT FREQUENCY, $F(j)$. LOGARITHMIC COORDINATES

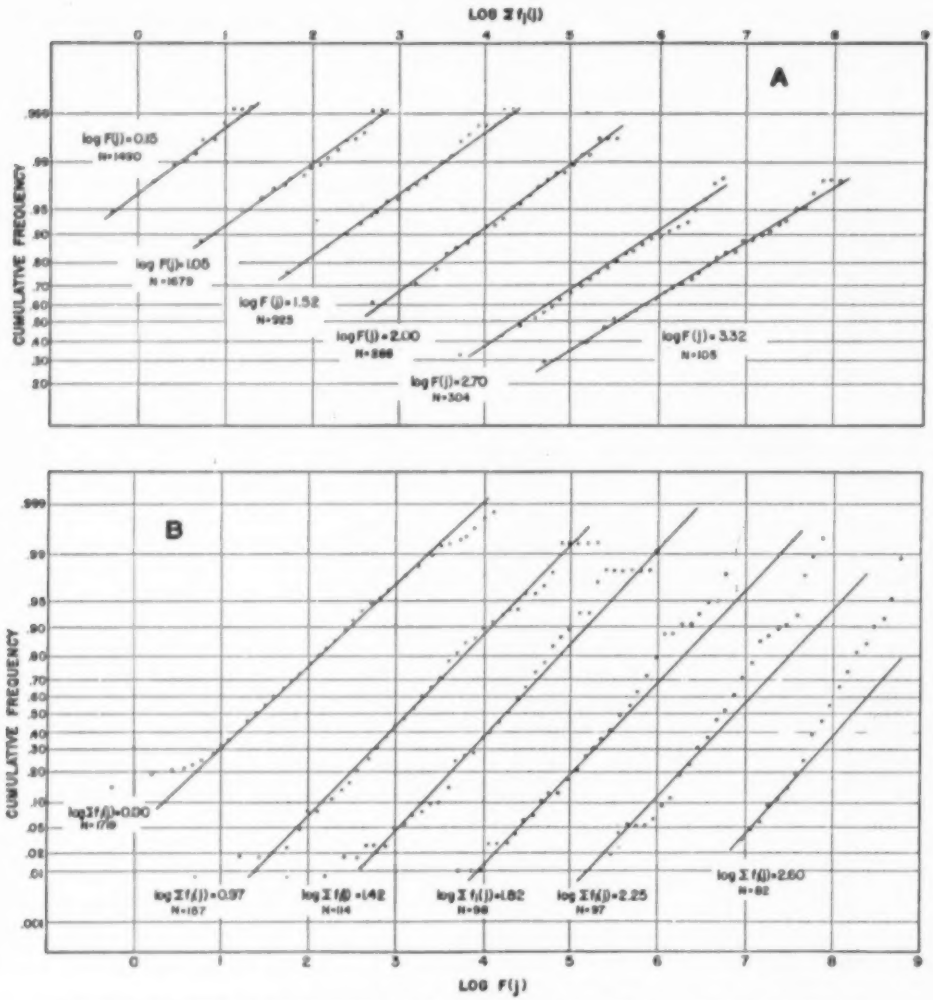


FIG. 2. A. CONDITIONAL DISTRIBUTIONS OF SUMMED KENT-ROSANOFF FREQUENCY FOR SELECTED VALUES OF MAGAZINE-COUNT FREQUENCY. B. CONDITIONAL DISTRIBUTIONS OF MAGAZINE-COUNT FREQUENCY FOR SELECTED VALUES OF SUMMED KENT-ROSANOFF FREQUENCY

Ordinate is proportional to the normal deviate, abscissa is logarithmic. The distributions have been separated by displacing each one to the right one log unit.

published as one of four word counts in the Thorndike-Lorge tables (8). These tables do not list the frequency of a word in each count if the combined frequency of the word in all four counts is less than 18. Moreover, the combined frequency itself is given only if it is greater than 3. In the data of Fig. 1 only magazine-count frequencies specifically listed in the tables are used. Words with combined frequencies of less than 18 are thus assigned a magazine-count frequency of zero in Fig. 1 even though they may actually have occurred up to 17 times in the magazine count.

The number of these words actually having a given magazine-count frequency can be estimated approximately by assuming that the occurrences making up the combined frequency of a word are distributed randomly

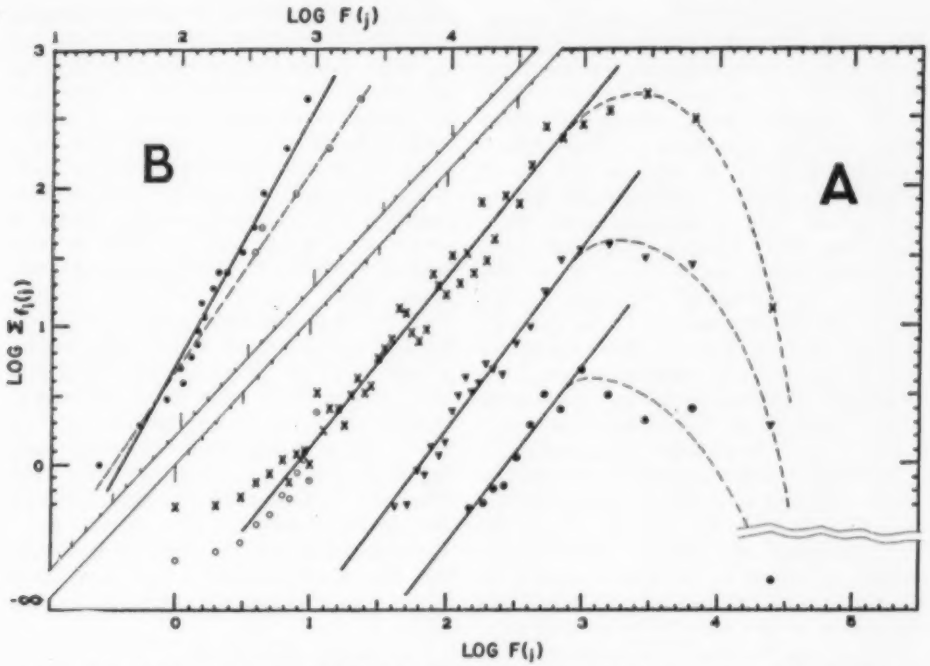


FIG. 3. A. REGRESSION OF SUMMED KENT-ROSANOFF FREQUENCY ON MAGAZINE-COUNT FREQUENCY. B. REGRESSION OF MAGAZINE-COUNT FREQUENCY ON SUMMED KENT-ROSANOFF FREQUENCY

Logarithmic coordinates. Solid circles, triangles, and crosses represent 50th, 75th, and 95th percentiles, respectively. Open circles indicate corrected data as described in text.

among the four equal component counts.[8] The summed Kent-Rosanoff frequency can be found for each word with a known combined frequency and assigned on a proportionate basis to the estimated magazine-count frequencies. In this way the words omitted from the published tables can be assigned values of $F(j)$ and $\Sigma f_i(j)$. These data can be used to correct Fig. 1.

The 95th percentiles of the $\Sigma f_i(j)$ distributions for these corrected data are shown by the open circles in Fig. 3-A (no correction is indicated for magazine-count frequencies of 12–17 where the change was not appreciable). The corrected data fall quite close to the theoretical straight line except for magazine-count frequencies of three or less (log $F(j) \leq 0.5$). In that range the correction is only partial because the published tables do not report words with combined frequencies of less than 4. The partial correction has a large enough effect, however, to indicate that complete correction would eliminate the remaining deviations of these points from the rectilinear regression line.

Variation of base probability with base associative probability constant. Consider next the con-

[8] Let Q be the combined frequency of a word in all four counts and q_1, q_2, q_3, and q_4 the unknown component frequencies, any one of which may be chosen to represent the magazine-count frequency. Then $\Sigma_i q_i = Q$ by definition and the limits $4 \leq Q \leq 17$ and $0 \leq q_i \leq 17$ follow from the nature of the Thorndike-Lorge tables. The number of possible arrangements of the four component frequencies that satisfy these conditions for given Q may be denoted by A_Q, and the number of these arrangements that assume a frequency m in the component representing the magazine count, $A_{m|Q}$. The probability that a word with a combined frequency Q has a magazine-count frequency m may be estimated from the ratio $A_{m|Q}/A_Q$. Suppose N_Q words are listed with combined frequencies of Q in the Thorndike-Lorge tables. Then $n_{m|Q} = N_Q(A_{m|Q}/A_Q)$ estimates the number of words with combined frequencies of Q that have magazine-count frequencies of m. The estimated number of words with magazine-count frequency m that are omitted from the published Thorndike-Lorge tables is therefore $n_m = \Sigma_Q n_{m|Q} = \Sigma_Q N_Q(A_{m|Q}/A_Q)$.

ditional distributions in the horizontal dimension of Fig. 1, which show the proportion of words with a given frequency in the Kent-Rosanoff tables that occur 0, 1, 2, ··· times in the magazine count. Figure 2-B shows several of these distributions on logarithmic-normal coordinates. In this case the sets of points do not describe parallel straight lines as in Fig. 2-A. Two artifacts, however, affect these data: the omission of words with magazine-count frequencies of less than 18, described in the preceding paragraphs; and the reversal of the regression line for log $\Sigma f_s(j)$ on log $F(j)$, shown in Fig. 3-A. When the effects of these two factors are taken into account, the theoretical distributions of log base probability, indicated by the straight lines in Fig. 2-B, turn out to be normal and homoscedastic.

The first factor is responsible for the leveling-off at the lower ends of the distributions of Fig. 2-B (the effect does not show up in the two right-hand distributions because they contain no points in the affected range). It has been pointed out that words with combined frequencies of less than 18 in the four word counts of the Thorndike-Lorge tables had to be assigned a frequency of zero in Fig. 1 although their true magazine-count frequencies may be as high as 17. This produces an excess of words in the interval $F(j) \leq 0.5$ and a corresponding deficiency in the interval $0.5 < F(j) \leq 17.5$. In the cumulative plot of Fig. 2-B, therefore, the points for $F(j) = 0.5$ (-0.3 on the logarithmic abscissa) lie well above their true values, while those for $F(j) = 17.5$ ($+1.2$ on the logarithmic abscissa) are correct. The correction procedure described in connection with Fig. 3-A eliminates this type of deviation altogether.

The second factor, the reversal of the regression line of Fig. 3-A, is responsible for the deviations at the upper ends of the distributions of Fig. 2-B. The effect of the reversal is to lower the positions of points at the extreme right of Fig. 1 from the positions they would occupy if the regression line of Fig. 3-A were rectilinear over its entire range. At log $F(j) = 3.0$ the median point is lowered about 0.1 log unit; at log $F(j) = 3.5$, about 1 log unit; and at log $F(j) = 4.0$, about 2 log units. Consequently, a horizontal slice through the upper part of Fig. 1 has a large proportion of its points in the range log $F(j) \geq 3.0$ removed by the reversal of the regression line. A distribution of log $F(i)$ at this level therefore has the form of a truncated normal distribution (2, §6.9). Distributions of log $F(j)$ taken at successively lower levels of Fig. 1 have proportionately fewer points removed by the reversal. The degree of truncation should decrease accordingly. At the very lowest levels of Fig. 1, in fact, the reversal actually adds more points from higher levels than it removes to still lower levels. The net effect at these levels is thus to add a small number of points in the affected range.

In Fig. 2-B, distributions of log $F(j)$ for successively lower levels of log $\Sigma f_s(j)$ are arranged from right to left. All six sets of data can be brought into accord with the theoretical straight lines by replotting them for the following assumed degrees of truncation (reading from right to left): 55%, 35%, 18%, 9%, 0%, and -10% (negative truncation indicates the addition of points instead of their removal). For each set of data the degree of truncation that gives the best fit to the theoretical line is of the order of magnitude required by the data of Fig. 1.

The medians of both empirical and theoretical distributions are used in Fig. 3-B to indicate the regression of magazine-count frequency on summed Kent-Rosanoff frequency. Solid circles represent medians of the empirical (truncated) distributions, open circles the medians of the theoretical (normal and homoscedastic) distributions. (For log $\Sigma f_s(j) \leq 1.3$ the solid circles represent both sets of medians.) Each set of data is well described by a straight line, the slope for the empirical medians (solid line in Fig. 3-B) being slightly steeper than the slope for the theoretical medians (dashed line).

Since the medians of the theoretical distributions have to be estimated from the theoretical lines of Fig. 2-B, their accuracy is open to question. As a test of their validity, the slope of the line fitted to the theoretical medians can be estimated by a second procedure. The truncation of the empirical distributions of log $F(j)$ does not extend to the lowest fractiles. Hence a line passed through a low fractile of the empirical distributions should have the same slope as the dashed line of Fig. 3-B representing the medians of the theoretical distributions.[2] The line of best fit to the 25th percentiles of the empirical distributions has a slope of 1.45 ± 0.1, compared to 1.40 ± 0.1 for the dashed line of Fig. 3-B.

By way of comparison, the dashed line of Fig. 3-B, representing the medians of the empirical distributions, has a slope of 1.85 ± 0.1.

Correlation between base probability and base associative probability. Only a small proportion of the words represented in Fig. 1—perhaps 0.5%—are involved in the reversal of the regression line in Fig. 3-A. The behavior of all other words is described by the rectilinear regression lines of Fig. 3-A and 3-B (dashed line). For these data, which satisfy the assumptions of a normal correlation surface (2, p. 585), the product-moment correlation coefficient ρ can be used to measure the degree of association between base probability and base associative probability. Numerical computation of ρ is impossible because of the presence of infinite scores. The correlation can be determined from the angle θ between the two regression lines, however, since

$$\tan \theta = \frac{(1 - \rho^2)\sigma_1\sigma_2}{\rho(\sigma_1^2 + \sigma_2^2)}.$$

(2, Equation 19.7.12). For the present data, θ is 3.5 degrees (Figs. 3-A and 3-B), σ_1 is 1.10 log units (Fig. 2-A), and σ_2 is 0.75 log unit (Fig. 2-B). The corresponding estimate of ρ is $+0.937 \pm 0.001$.

This correlation must be interpreted in the light of several adverse factors affecting the data: the difference of 30 years between the dates of the two sets of data; the different composition of the two populations of subjects; the contrasts between the written language of the magazine count and the spoken language of the Kent-Rosanoff experiment; the small number of stimulus words tested in the Kent-Rosanoff study; and the use of the approximation to base associative probability instead of its exact definition. All these factors can be expected to introduce uncorrelated variance into the results. The true correlation between base probability and base associative probability, uncontaminated by these sources of error, must therefore be close to 1.0. Stochastic equivalence thus appears to apply to the words considered in this section.

The reversal of regression at high magazine-count frequencies. The words having the highest magazine-count frequencies are the ones that are responsible for the inversion in the regression line of Fig. 3-A. Table 1 lists the 25 words having the highest magazine-count frequencies along with their summed Kent-Rosanoff frequencies. Without exception, the high magazine-count frequencies of these words are matched with very low summed Kent-Rosanoff frequencies. The average summed Kent-Rosanoff frequency is only 0.2% of the average magazine-count frequency, after correction for the difference in size of sample.

TABLE 1

SUMMED ASSOCIATIVE FREQUENCIES FOR THE 25 WORD CLASSES WITH THE HIGHEST MAGAZINE-COUNT FREQUENCIES

Word	$F(j)$	$\Sigma f_i(j)$	
	Magazine Count N = 4,591,125	Kent-Rosanoff N = 100,000	Jenkins N = 19,100
1. the	236472	0	28
2. and	138672	0	54
3. a	117222	0	12
4. to	115358	0	126
5. of	112601	0	38
6. I	89489	6	84
7. in	75253	0	29
8. was	58732	0	44
9. that	55667	0	73
10. it	52107	0	196
11. he	49268	0	82
12. you	42581	13	219
13. for	39363	6	54
14. had	34341	0	42
15. is	33404	0	135
16. with	32903	0	42
17. her	31824	1	60
18. she	31087	0	70
19. his	30748	0	104
20. as	30693	0	10
21. on	30224	1	64
22. at	26250	0	61
23. have	24456	19	47
24. but	23704	0	27
25. me	23364	14	242
Sum	1,535,783	54	1943
Sum/N	0.33451	0.00054	0.10171

Study of Table 1 discloses that all the words listed there are of a characteristic type. They are words whose primary function is to connect other words in continuous discourse: articles, conjunctions, auxiliaries, prepositions, pronouns. These may be called *interstitial* words. The low frequencies of such words in the Kent-Rosanoff tables could be expected on two grounds. First, the subjects in the association experiment are instructed to emit only one-word utterances. The absence of continuous discourse eliminates the principal function of interstitial words, thereby reducing the number of occasions on which they are called forth. Second, it has been well established that sub-

jects tend to give associations of the same part of speech as the stimulus words (10, p. 347). The complete absence of interstitials among the 100 Kent-Rosanoff stimulus words would therefore result in low frequencies for those words as associations. The former interpretation of the result implies that interstitial words represent a special type of word to which the property of stochastic equivalence does not apply, while the latter interpretation holds that the exceptional behavior of these words is merely an artifact of the particular association experiment that has been analyzed.

The best evidence on the question is found in a recent association experiment by Jenkins[4] in which over 50 interstitials, including all but three of the words in Table 1 (*in*, *her*, *as*), were among the 200 stimulus words. Table 1 shows the frequencies, summed over all 200 stimulus words, with which the words listed there occurred as responses in the Jenkins study. The effect is striking: there is a 200 to 1 increase in frequency over the Kent-Rosanoff experiment. The 25 words of Table 1, while they constitute only 0.05% of the Kent-Rosanoff responses, make up 10% of the Jenkins responses and 33% of the magazine count. Inclusion of interstitial words in the list of stimulus words thus goes a long way toward closing the gap between the summed associative frequencies and the magazine-count frequencies of interstitial words. Use of the complete formula for base associative probability instead of the approximation would narrow the gap even more.

The Jenkins experiment, based on fewer than 100 subjects, is not extensive enough to warrant a more elaborate analysis. But it shows clearly that the particular selection of stimulus words used by Kent and Rosanoff is chiefly responsible for the reversal of the regression line in Fig. 3-A. The sharpness of the contrast between the Jenkins results and those of Kent and Rosanoff further suggests the possibility that stochastic equivalence may apply to interstitial words as well as to noninterstitial ones. A large and carefully designed experiment would be required, however, in order to test this possibility adequately.

[4] I wish to thank Dr. J. J. Jenkins of the University of Minnesota for generously placing his data at my disposal and Miss Evelyn Segal for preparing the computations.

SUMMARY AND CONCLUSIONS

The summed associative probability of a word, measured by its total frequency in the Kent-Rosanoff tables, is considered in relation to the probability of emission of the word in general linguistic usage, measured by its frequency in the Lorge magazine count. Two features stand out in the data: a high positive correlation (.94) for magazine-count frequencies of less than 800; and a sharp reversal in this relationship for higher magazine-count frequencies. Perhaps neither result, taken singly, is as striking as the fact that both appear so clearly in the same set of data.

The inverse relationship for high magazine-count frequencies is produced by a fairly small number of words of a special type, called *interstitial* words because their chief function is to connect other words in continuous discourse (i.e., conjunctions, articles, prepositions, auxiliaries, pronouns). Such words were systematically excluded from the list of stimulus words used by Kent and Rosanoff. As a consequence, they were rarely emitted as associations despite their high magazine-count frequencies. Data from another association experiment, in which interstitial stimulus-words were included, indicate that the reversal of correlation found in the Kent-Rosanoff data can largely, if not completely, be eliminated by an unbiased selection of stimulus words.

The high correlation found for all other words indicates that the average probability that a given word will be emitted as a response in the word-association experiment is the same as its probability in general discourse. The language of association can therefore be said to be *stochastically equivalent* to the language of general discourse. This relationship implies that a person's selection of words is not affected by the artificial laboratory situation required for the association experiment. Hence the effect of a given stimulus word upon a subject's response must be due to a redistribution of the set of factors governing emission of words in general discourse. The results thus serve to validate the association experiment as a method for studying under laboratory control the processes governing the emission of words in general discourse.

REFERENCES

1. COFER, C. N., & SHEVITZ, R. Word-association as a function of word-frequency. *Amer. J. Psychol.*, 1952, **65**, 75–79.
2. HALD, A. *Statistical theory with engineering applications*. New York: Wiley, 1952.
3. HOWES, D., & OSGOOD, C. E. On the combination of associative probabilities in linguistic contexts. *Amer. J. Psychol.*, 1954, **67**, 241–258.
4. KENT, G. H., & ROSANOFF, A. J. A study of association in insanity. *Amer. J. Insanity*, 1910, **67**, 37–96; 317–390.
5. O'NEIL, W. M. The effect of verbal association on tachistoscopic recognition. Univer. of Minnesota Stud. on Verbal Behavior, Tech. Rep. No. 4, 1953.
6. RUSSELL, W. A., & STORMS, L. H. Implicit verbal chaining in paired-associate learning. *J. exp. Psychol.*, 1955, **49**, 287–293.
7. SKINNER, B. F. The distribution of associated words. *Psychol. Rec.*, 1937, **1**, 71–76.
8. THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Teachers College, Columbia Univer., 1944.
9. THUMB, A., & MARBE, K. *Experimentelle Untersuchungen über die psychologischen Grundlagen der sprachlichen Analogiebildung*. Leipzig: Engelmann, 1901.
10. WOODWORTH, R. S. *Experimental psychology*. New York: Holt, 1938.
11. ZIPF, G. K. *Human behavior and the principle of least effort*. Cambridge, Mass.: Addison-Wesley, 1949.

Received March 12, 1956.

LOOKING AT OCCUPATIONS[1]

WILLA FREEMAN GRUNES

University of California, Berkeley[2]

How does the young American look at jobs? What groups of jobs appear in his psychological field? What do these groups look like to the perceiving individual? Do different social classes see these things differently? These are phenomenological questions, of the sort MacLeod (9) meant when he suggested that the phenomenological method, found so fruitful in approaching classical problems in perception, be applied also in the field of social psychology. They concern the social world which is psychologically "there" for the subject, quite apart from what the experts might say is, or ought to be, there.

There have been few studies that attempt to answer such questions about aspects of the social world by attempting to discover perceived units and their characteristics. Methods for such investigation are woefully lacking and difficult to invent. The present study is presented as suggestive of new ways for mapping perceived social worlds.

If we can determine what people see, we can better understand what they think and do. Specifically, in the realm of occupations, we can better see why they make their particular vocational choices, and why too many choose certain types of overcrowded fields while too few choose others.

Occupational perceptions were selected for this study partly because they seem important for revealing the psychological meaning of social class memberships. Such meaning has long been neglected in research on class in America. These perceptions also offer a meeting place between psychology and economics and hence offer a promising area for mutual enrichment of both fields.

METHOD

Sample

About 150 students in each of eight high schools were tested in 1952–53. Four Northern California schools were included: Oakland, McClymonds, San Lorenzo, and Hayward. There were also two Southeastern schools: Talledega in Alabama and Orangeburg in South Carolina; one Northwest Central school: Bismarck in North Dakota; and one Northeastern: Clifton Heights in Pennsylvania. The communities of the school vary widely in size, social class, race, and in degree and type of industrialization. The diversity seemed desirable in order to assess the generality of the findings.

Procedure

Two group-administered devices, the Grouping Test and the Pick-A-Job Test, were developed and pretested for purposes of revealing the way perceived occupations tend to be grouped by high school students; i.e., the perceived structure of the world of jobs. Both were indirectly inspired by various studies by Krech [(8) and recent unpublished work] on the way rats and also men structure the world around them. In order to make it easy and threatless for subjects to reveal their perceptions, both devices are disguised on purpose. The subject is provided with an incomplete situation, out of which he must make some sense of his own in order to deal with it. Such a disguised and unstructured approach seems appropriate to an area of content in which subjects may be unable or unwilling to verbalize what they perceive. Through these two devices, a perceptual structure was sought which could be regarded as common to most high school students. The assumption was nevertheless avoided that all social classes—or all individuals—possess a similar structure, or that any perceived structure necessarily reflects objective social reality. In this respect the method differs from that used by Lloyd Warner and his associates (10) to study class structure.

Grouping Test. On the Grouping Test, the subject was presented with a list of 51 varied occupations, selected to represent each of the major categories in the Dictionary of Occupational Titles (11). He was given the following written instructions (somewhat abridged here) along with some oral clarification:

> Let us see how fast you can think about various occupations. Group the following occupations as many ways as you can. For example, make groups according to what kinds of people work at the jobs, what kinds of work they do, etc. Give each group a title that tells what kinds of jobs belong under it. . . . At least two jobs must be in each group you make. You may use the same jobs under as many titles as you wish. For example: People who would probably know how to give

---

[1] The material discussed is adapted from a doctoral dissertation (6) done under the guidance of Professor Mason Haire. Especially extensive assistance was given by H. D. Mugass, R. R. Lange, W. C. Rasmussen, Dorothy W. Gottlieb, Katherine M. Cavanagh, Leon Festinger, and Camille J. Miller, as well as by the administrators and their staffs of the schools involved.

[2] The author now resides in Mandan, North Dakota.

first aid: airline stewardesses, doctors, nurses, physical education teachers. . . .

The more groups you make within the time limit (15 minutes) the higher your score. You can use any types of groups you can think of. . . . Be sure to include all of the occupations on the list that belong under each of your titles, because you also get a higher score for every job you list that belongs there. Here is the list of occupations. . . .

The resulting data were tabulated and coded to indicate the jobs which tend to be placed together in the same groups, the grouping categories reflected in the titles, and the jobs most frequently placed under each type of title. A set of job clusters common to most students' view of the world of occupations was discovered. Content analysis of the titles was done by two coders working independently, using only categories which yielded at least 80 per cent agreement.

Since completing the present study, the author has learned of a similar effort by Campbell (1) to develop a method for establishing subjective occupational groupings. His study failed, however, to reveal any common job clusters. The present method differs from his in permitting the subject to use each occupation in as many groups as he wished and to make as many (overlapping) groups as he could, as well as in the procedures used to analyze the data. These provisions appear more appropriate than Campbell's to the presumably complex nature of the perceived structure of the occupational world.

Pick-A-Job Test. The Pick-A-Job Test offered the subject brief descriptions of six different people and asked him to assume the role of vocational counsellor and to suggest five appropriate jobs for each one. For example, the first description appeared thus:

A. Man	1._____
good-looking	2._____
clean	3._____
well-educated	4._____
self-confident	5._____

Results were analyzed to indicate the typical sorts of jobs suggested for each description.

The personality descriptions used on the Pick-A-Job Test were derived from earlier use of what may be called the Haire Describe-A-Person method, a modification of a procedure originally developed by Haire and described by Haire and Grunes (7). High school students were given a few rather neutral facts about a fictitious person, and also his occupation, and were asked to guess the sort of person he was. Their guesses were coded for content and compared with descriptions obtained from similar subjects when the occupation was omitted from the given facts. Differences between the two sets of descriptions revealed stereotypes associated with the occupations selected for study. The Pick-A-Job Test was based on the typical descriptions thus obtained of businessman (Description 1 above), ditch-digger, actress, and farmer. Additional descriptions based on the typical businessman and ditch-digger were included, with the characteristics ascribed in each case to a woman.

Besides these procedures, the Gough Home Index (4), a measure of relative social status of high school students, was administered. It yields a continuous distribution of scores, but was here employed to classify subjects into three status groups (low, middle, and high) by cutting the distribution for each school at the two points where the large central frequencies began to fall off.

TABLE 1

SOCIAL CLASS DIFFERENCES IN FREQUENCY OF GROUPING OTHER OCCUPATIONS WITH DOCTOR

(Entries represent percentage of groups involving "doctor" made by each social class)

Co-Member	Home-Index Social Class		
	Lower	Middle	Upper
Nurse	31	66	53
College Professor	38	59	80
Teacher	44	56	73
Social Worker	13	24	47
Air Line Stewardess	56	38	33

RESULTS

Deriving the Job Clusters

Data were compiled on each job listed on the Grouping Test, showing, for each social class, the other jobs with which a given job was grouped and the types of titles under which it was listed.[3] As a sample, the following information was gathered about the "doctor":

In at least 50 per cent of the instances in which doctor was listed, the following jobs appeared in the same group: nurse, college professor, and teacher. In at least 40 per cent, the engineer and minister were in the same group.

The dish-washer, ditch-digger, farmer's wife, hired man on farm, rancher, and janitor were in the same group with doctor less than 5 per cent of the times it appeared.

Doctor appeared in at least 50 per cent of the groups in which each of the following jobs appeared: nurse, college professor, teacher, engineer, social worker, reporter, businessman, and business-woman.

Table 1 shows the percentage of groups made by each social class that included both the doctor and each of the five occupations listed. Note that the higher status subjects tended to group doctor with other professionals more consistently than did the lower status subjects.

The categories of Grouping Test titles under which doctor was listed at least 50 per cent of the times they occurred include: educational barriers, much education, good looking, lot of money, personality barriers, social skills, ability barriers, brains, verbal skills, "skilled," "trained," training barriers.

Doctor was listed under less than 50 per cent of the titles categorized as follows: charm, strong, physical barriers, little or no education, "common labor," dumb, not own boss, little money.

[3] All the data are available in tabular form in the author's dissertation (6).

TABLE 2
A SAMPLE JOB CLUSTER MATRIX: CLUSTER A

	Job Title									
	Doc.	Nur.	Coll.	Teach.	Eng.	Min.	Social	News.	Busm.	Busw.
Co-Members										
Doctor	X	++	++	++	++	++	++	++	++	++
Nurse	++	X		++	+	++	++	+		+
College Professor	++	++	X	++	++	++	+	++	++	++
Teacher	++	++	++	X	++	++	++	++	+	++
Engineer	+				X	+				
Minister	+	++	++	+	+	X	++	+		+
Social Worker					+	+	X			
News Reporter						+	+	X		+
Businessman			+	+		+	+	+	X	++
Business Woman			+			+	+	+	++	X
Air Line Stewardess		+				+	++	+		+
Phys. Ed. Teacher					+	+		etc.*	etc.*	etc.*
Nonmembers										
Dish Washer	–		–	–	–	–		–	–	–
Ditch Digger	–	–	–	–		–			–	–
Farmer's Wife	–		–	–	–	–		–	–	–
Hired Man, Farm	–		–	–		–			–	–
Rancher	–		–	–		–			–	–
Married Woman			–	–	–			–	–	–
Welder			–			–			–	–
Truck Driver			–	–					–	–
Maid			–		–				–	
Housewife			–		–				–	
Janitor	–		–						–	
Garbage Collector			–						–	
Factory Worker			–			–				
Farmer			–							
Cleaning Woman			–							

Note—++ means that 50 per cent or more of the times that the occupation at the top of the column was listed, it appeared in the same group as the occupation at the left of the row.

+ means that such co-grouping occurred in at least 40 per cent, but not as many as 50 per cent of the total listings of the occupation at the top.

– means that the occupation at the left appeared in fewer than 5 per cent of the groups in which the occupation at the top occurred.

* In the case of news reporter, businessman, and business-woman, associations were found with additional co-members not listed here. Data for the other occupations are complete.

From the above data, it was predicted that the first Pick-A-Job description, as given above, would elicit the suggestion of doctor, among others. Actually, although the students had the whole world of occupations from which to choose, and five different choices to make, doctor accounted for fully 10 per cent of all jobs chosen for that description.

Similar data for each job were combined to define and characterize the commonly perceived job clusters. Table 2 gives, for example, the matrix of relationships which led to the definition of the cluster (Cluster A) to which the doctor belongs. The data presented also provide some evidence that Cluster A overlaps with another cluster.

A complete matrix, similar to Table 2 but including all 51 jobs on the Grouping Test, was used in determining the commonly perceived clusters.[4] Unless an occupation was grouped with at least one member of a cluster at least 40 per cent of the time, it was not considered to be even a borderline member of that cluster. To be considered a clear member, it had to be grouped at least 50 per cent of the time with most of the other clear members of the cluster.

A number of jobs were found to be borderline or clear members of two clusters, and these jobs were taken as evidence that the clusters overlap. No jobs were found to be even borderline members of more than two clusters.

Confirmation of the clusters thus derived was found in the "nonmembers," which always came from "distant" (nonoverlapping) clusters, and in the data from the Pick-A-Job Test. Students tended to choose jobs from a highly restricted set of clusters for each description on the Pick-A-Job Test. These clusters were found to be predictable from a knowledge of the cluster membership of the job the stereotype of which had been drawn upon in formulating the description. For example, Description 1 (given

[4] The matrix composes Part IV of the Appendix of the dissertation, pp. 307–316 (6).

above) elicited almost exclusiveiy members of the cluster (A-B) to which "businessman" belongs and members of the two clusters (A and B) with which A-B overlaps.

Implicit criteria in grouping. Characteristics of the clusters were determined by an analysis of the type of titles under which the cluster members were most frequently grouped—or not grouped—on the Grouping Test. As shown by the types of titles they choose for their groups, students typically show consciousness of certain dimensions in thinking about jobs, most notably: barriers to attaining the jobs (especially educational barriers), status, skill and knowledge, and the nature of the work.

The empirical job clusters. The overlapping job clusters and the outstanding characteristics ascribed to them, as revealed by the Grouping Test and confirmed by the Pick-A-Job procedure, are as follows:

Cluster A. These jobs are characterized as requiring much education, ability, brains, social and verbal skill, and as yielding much money; they are almost never seen as requiring physical strength. Included from the list provided are college professor, minister, doctor, nurse, teacher, and less clearly engineer, social worker, reporter, and businessman.

Cluster A-B. These jobs have characteristics similar to those of Cluster A, but with much less stress on educational barriers. They include businessman (also a member of A), and business-woman (also a member of B).

Cluster B. These jobs are seen as "business" and as requiring verbal skills. They include business-woman, office-worker, secretary, stenographer, bank clerk, and, less clearly, salesman, saleswoman, and store clerk.

Cluster C. These jobs are seen as divided into two subclusters: C-1, characterized by skill and strength, and C-2, seen as hard work and "common labor" done by a worker who is strong and not his own boss. C-1 includes carpenter, mechanic, and welder; C-2 the farmer, rancher, hired man on a farm, truck-driver, and less clearly, ditch-digger.

Cluster C-D. These jobs are seen as involving hard work, "common labor," and as requiring little education or anything else except strength. The workers are seen as stupid and as not their own bosses. They include ditch-digger (also a member of C), garbage-collector, and janitor.

Cluster B-D. These jobs are seen as "business" and as requiring social skill. They include salespeople and store clerks (also in Cluster B) and also waiters and waitresses.

Cluster D. These jobs are seen as requiring little (not even strength), as "common labor," and hard work, and their holders as stupid people of little education who are not their own bosses. They include dishwasher, cook, maid, housewife, married woman who stays home and keeps house, cleaning woman and, less clearly, elevator operator, waitress (also a member of B-D), and janitor (also a member of C-D).

Certain occupations on the list could not be placed in the above system of interrelated clusters, including one rather weak group including actresses, airline stewardesses, and models, and a strong one including only policemen and firemen. Jobs which seemed to stand completely alone were factory worker (distantly related to C-D and D), aviator, musician, physical education teacher, farmer's wife, tailor, barber, foreman, railroad conductor, and baseball player. A still more exhaustive list of occupations might well reveal further relationships.

Prestige Hierarchy

In the characteristics ascribed to the clusters, the prestige hierarchy inherent in the common cognitive structure is revealed, with positive and barrier-type characteristics both tending to be concentrated in the A and A-B Clusters, and least in Cluster D and the clusters overlapping it. Physical strength is an interesting exception to this rule. This prestige hierarchy agrees well with the conclusions of other investigators (2, 3) who have asked various samples of Americans to rank lists of jobs according to their "social standing."

There is a large gap between the perception of professionals, business, and office people (Clusters A, A-B, and B, but not B-D) on the one hand, and manual and mechanical workers on the other, with only relatively rare cases occurring in which they are seen as members of the same group. Some of the evidence for this deep gap may be seen by studying the "non-members" in Table 2. All of these cases of extremely rare grouping with Cluster A occupations are from Clusters C, D, and C-D. The gaps between skilled and unskilled workers and between clerical and professional workers are not nearly so wide.

Stereotyped Job Choices

Students agree remarkably well on the occupations and clusters of occupations seen as appropriate for the people described on the Pick-A-Job Test. For each description, their choices are sharply limited to a few salient possibilities out of the thousands of occupations from which they were free to choose. In answer to Description 1, for example, the following job categories accounted for about 70 per cent of all the suggestions: salesman, manager, businessman, teacher, doctor, lawyer, and "office worker."

TABLE 3

OCCUPATIONS MOST COMMONLY GROUPED WITH COLLEGE PROFESSOR ON THE GROUPING TEST BY SOCIAL CLASS

Co-Member	Cluster	Percentage of groups which contain both the occupation listed and college professor		
		Lower Class	Middle Class	Upper Class
Doctor	A	53%	70%	75%
Minister	A	33	53	63
Teacher	A	73	68	81
Engineer	A	30	44	58
Businessman	A-B	50	47	44
Business-woman	A-B	60	43	42
Office worker	B	33	26	17

Note.—Few of the differences in percentage shown are statistically significant, but they seem important because of their consistency. Those for college professor are typical of those for all other Cluster A occupations. There were no relevant cases showing an opposite trend.

For Description 2—a man who is hard-working, strong, not very bright, and likes to follow orders—students mainly choose jobs from Clusters C and C-D, which contain mostly masculine occupations. The woman with the same traits is steered mainly toward Clusters B-D and D. For the other descriptions, however, the sex of the person described changes the particular jobs selected, but not the clusters from which they come.

Group Differences in Perception of Occupations

Class differences. Although there is considerable agreement among students about the common pattern of clusters, there are some social class differences in cognition of the occupational world. Most notably, students from the upper classes tend to make a sharper distinction between business and professional people than do the lower classes. The latter tend to lump Clusters A and A-B and B into one high-prestige cluster. They also tend to exclude from that cluster certain occupations often placed there by upper classes—such as the social worker, minister, engineer, reporter, and musician. Table 3 presents a sample of the evidence for these conclusions, using the occupations grouped with college professor on the Grouping Test as examples. Each value in the table represents the percentage of the total number of groups, made by the subjects in a particular social class, which contain both the occupation at the left and "college professor." Note that the college professor is grouped with other professionals more consistently, and with business people less frequently, as status rises. (See also the data on doctor listed above.) In general, the upper classes seem to be more finely differentiated about the upper occupational strata. There is also some less clear evidence that the lower classes are more differentiated about the lower prestige jobs.

The class differences in characteristics attributed to particular jobs follow lines predictable from the class differences in the job clusters themselves. Subjects from the lower classes tend more often to include jobs from Clusters A-B and B—as well as A, which all classes stress—when listing jobs that share certain positive and barrier-type characteristics, especially those concerning education and ability.

Certain particular occupations on the Grouping Test list were placed in different clusters by different classes. The clearest example is the "engineer," interpreted as a professional by upper classes and as a skilled mechanical worker by lower.

These class differences are evident in coherent patterns in the data, but could not feasibly be subjected to statistical test. The group differences described below, except where otherwise noted, were found to be significant at the .05 level of confidence.

Subjects higher in social status appear to be relatively lacking in both status and barrier consciousness, but relatively strongly conscious of skill and knowledge, pace of work, and money. It may be that students are less aware of the job aspects about which they feel less insecure. The fact that students whose breadwinners are in manual or mechanical work are relatively less conscious of strength seems to support this inference.

The Pick-A-Job Test results reflect the greater tendency of the lower classes not to differentiate the top three job clusters. While higher status subjects restrict themselves closely to professionals and business people for both the male and female version of Descrpition 1, clerical occupations are often selected by lower classes. For example, clerical work is suggested for the male version by 31.3 per cent of the lower-class subjects, by 15 per cent of the middle class, and by only 2.6 per cent of the upper class (difference significant at the .01 level). For Description 2, upper classes tend to

name more stereotypically low-prestige jobs like the garbage-collector and ditch-digger—perhaps again indicating their coarser distinctions among lower-strata occupations.

Sex differences. The study revealed some outstanding differences in the way boys and girls see the world of jobs. Through the titles they used on the Grouping Test, girls showed more awareness—or at least more articulateness—concerning all the common types of job aspects, with the exceptions of security, money, and hours of work. If the female is more status conscious, as has been claimed, she is also more conscious of a large number of other perceptual dimensions of jobs.

Sex differences were also found in the answers to Description 1 and its feminine counterpart on the Pick-A-Job Test, with the girls showing a stronger tendency than boys to name teaching jobs for both versions. Boys tend to select clerical rather than professional jobs for the female version, while girls choose mostly professional jobs, as do both sexes for the male version—perhaps a sex difference in willingness to limit capable females to lower social position.

Regional differences. In broad outlines, students in all eight schools see the same world of occupations. But consistent trends in the data indicate that those in the large industrialized communities tend, like higher status subjects, to see a relatively sharp professional business distinction. Students in the Southeastern schools appear, through their Grouping Test titles, to be more conscious of status and education than the others tested; but otherwise there is considerable regional similarity in the types and relative potencies of dimensions used for categorizing occupations. The social worker is the outstanding example of a job that is grouped differently in different regions: it is much more clearly professional in the large Northern cities.

Despite the class, sex, and regional differences mentioned, in general the students seem to agree rather closely in their basic perceptual structure concerning occupations.

DISCUSSION

The Grouping Test technique, in which a collection of items is set before the subject for him to record the groups or units that he perceives spontaneously, is easily adaptable to many problems of social perception. Instead of occupations, one could list, for instance, names of political figures, nationalities, religions, brands of food, business corporations, or periodicals. The general method of the Pick-A-Job Test is likewise adaptable. Since it involves giving the subject hypothetical group characteristics and asking him to name members of the group, it could be called the Pick-The-Members technique. In other adaptations, subjects could be asked to name, instead of jobs, clubs to which the described individual might belong, public policies he might be likely to favor, movies he might enjoy, etc.

The Haire method (Describe-The-Person or Describe-The-Member), used here in conjunction with the Pick-A-Job Test, seems generally adaptable for yielding hypotheses about the characteristics of perceived groups. Together the two methods can be used to confirm Grouping Test data about the composition and characteristics of the perceived units. If the conclusions from the Grouping Test are valid, use of the Pick-The-Members technique on the same subjects should yield the members of the appropriate Grouping Test cluster, when that cluster's characteristics are used to form the description on the Pick-The-Members Test. And the Describe-The-Person method should tend to yield the characteristics found through the Grouping Test for the group of which the Person is seen as a member. The various methods were used in this manner in the present study to complement and validate each other.

Some of the implications of the study for vocational guidance are discussed elsewhere (5).

SUMMARY

A phenomenological study of the way American high school students perceive occupations is presented. A perceptual structure consisting mainly of seven overlapping job clusters is found to be characteristic of most subjects. Class and regional differences in the structure are analyzed and the different attributes ascribed to each cluster are examined. Two original types of methods are used which are adaptable to many other problems in social perception.

REFERENCES

1. CAMPBELL, J. D. Subjective aspects of occupational status. Unpublished doctor's dissertation, Harvard Univer., 1952.

2. COUNTS, G. S. Social status of occupations. *Scholast. Rev.*, 1925, **33,** 16–27.
3. DEEG, M. E., & PATERSON, D. G. Changes in social status of occupations. *Occupations*, 1947, **25,** 205–208.
4. GOUGH, H. A short social status inventory. *J. educ. Psychol.*, 1949, **40,** 52–56.
5. GRUNES, WILLA FREEMAN. On perception of occupations. *Personnel Guid. J.*, 1956, **34,** 276–279.
6. GRUNES, WILLA FREEMAN. The American high school student looks at the world of occupations. Unpublished doctor's dissertation, Univer. of California, Berkeley, 1954.
7. HAIRE, M., & GRUNES, WILLA FREEMAN. Perceptual defenses: processes protecting an organized perception of another personality. *Hum. Relat.*, 1950, **3,** 403–412.
8. KRECHEVSKY, I. Hypotheses in rats. *Psychol. Rev.*, 1932, **6,** 516–532.
9. MACLEOD, R. B. The phenomenological approach to social psychology. *Psychol. Rev.*, 1947, **54,** 193–210.
10. WARNER, L., MEEKER, MARCHIA, & EELLS, K. *Social class in America.* Chicago: Science Research Associates, 1949.
11. Dictionary of occupational titles. (2nd ed.) Washington, D. C.: U. S. Bureau of Employment Security, 1949.

Received March 13, 1956

AGE DIFFERENCES IN THE PERCEPTION OF CLOSURE[1]

HAROLD BASOWITZ AND SHELDON J. KORCHIN

Institute for Psychosomatic and Psychiatric Research and Training, Michael Reese Hospital

THE changes associated with aging are many and varied. Certain alterations in sensory acuity, speed of response, and performance on motor, memory, and selected intellectual tasks have been reasonably well established both by common observation and controlled experimentation (6, 8, 17). Yet the number of investigations concerned with more complex perceptual functioning in older persons has been surprisingly small. Of particular interest are comparative studies of Rorschach protocols typifying groups of differing ages and their implications for distinctive aspects of personality organization (3, 4, 5, 11, 18).

In the course of a larger research project dealing with psychological and neurophysiological correlates in aging, some provocative findings emerged related to a more circumscribed type of perceptual ability—that of closure. It is the purpose of this paper to consider this material. More specifically, the data to be presented involve the performance of two groups widely separated in age on two kinds of perceptual tasks which earlier factorial studies have claimed to reflect "the ability to fuse a perceptual field into a single percept" (21) and "flexibility of closure" (20). Within a broader context, this study is conceived as an attempt to enhance our general understanding of cognitive functioning in senescence. Related investigations are reported separately (14, 15).

METHOD

Two populations were sampled: a "young" group of 16 doctors and nurses composed of 5 men and 11 women, and an "old" group of 16 residents, 8 men and 8 women, in a home for the aged. The ages of the young ranged from 22 to 33, with a mean of 26.8 years; those of the old from 68 to 88, with a mean of 78.1 years. All *Ss* were physically and mentally healthy and had adequate vision, using corrective lenses where necessary. Since one of the techniques to be discussed, Concealed Figures, has been found by other workers to be saturated with an intellectual factor (20), the groups were chosen so as to be roughly equated for intelligence. They were drawn from 24 young and 36 old persons used in the total study, to whom a battery of tests had been administered which included the Vocabulary subtest of the Wechsler-Bellevue Intelligence Scale (22) as well as the two perceptual tasks of concern in this paper. Matching for general intelligence was performed independently on the basis of vocabulary level, since it has been adequately demonstrated that this index declines relatively little in the normal process of aging (1, 2, 10, 22). The mean and *SD* of unweighted vocabulary scores for the young *Ss* so selected were 26.8 and 4.6; for the old, 28.6 and 4.8, respectively. Although these restricted groups enable us to focus most directly on influences of age uncontaminated by differences in intellectual ability, parallel analyses of the data have also been carried out for the larger parent populations. In all instances the age differences in performance are similar in direction and extent to those of the present study, at times being perhaps more extreme.

Two procedures were used:

Gestalt Completion Test. In this task, adapted by Thurstone (20) from a technique developed by Street (19), *S* is presented with a series of 72 ambiguous drawings which he is required to identify. Portions of the drawings are obliterated and at first view the material appears to consist of diverse and unrelated black blotches. Actually the fragments are arranged so as to suggest the outlines of distinct persons, objects, and/or situations. The accomplishment of closure involves organizing the seemingly scattered elements so that a compelling and unified whole emerges. In his factor-analytic studies of mechanical aptitude Thurstone (21) concluded that the test represented "the ability to fuse a perceptual field into a single percept."

Following some examples, the instructions were: "When the signal is given, turn the page and identify some more pictures of the same kind. Work as rapidly as you can. *If some pictures are too difficult, skip them, and return to them later if you have time.*" Scoring criteria corresponded to those followed by Campbell and Zaks[2] in recent unpublished studies. Responses were scored 1 when the activities depicted or all of the persons and objects in the picture were accurately described. Half credit was allowed when partial integrations occurred and these were essentially correct.

Concealed Figures. Thurstone's (20) version of the Gottschaldt Figures Test (9) was employed. Forty-nine rows of figures are given, with a simple geometric design appearing at the left of each row followed by four more complex designs. The *S* identifies each of the more complex designs in which the simple figure is em-

[1] This study was supported by the U. S. Army under Contract No. DA-49-007-MD-261 through the Medical Research and Development Board, Office of the Surgeon General, Department of the Army, Washington, D. C. We gratefully acknowledge the aid of Mr. Ben Grossman, Director, and the staff of the Drexel Home, Chicago, in providing subjects and facilities. Dr. Albert Caron and Mrs. Ann Gunther were responsible for much of the data collection and Dr. Helen Heath for the statistical analyses.

[2] Personal communication.

TABLE 1
MEAN PERFORMANCES OF YOUNG AND OLD GROUPS ON GESTALT COMPLETION TEST

Group	*N*	Attempted	Right	Wrong	Omitted
Young	16	63.1	44.7	6.0	12.4
Old	16	60.7	15.5	15.9	29.4

bedded. Since the complex figure has "good gestalt" its properties of closure must be resisted as *S* searches for the smaller figure potentially hidden therein. The principal component of this test was termed by Thurstone (20) "flexibility of closure."

In essence the instructions stated: "In every row of designs the *figure* is concealed in at least one drawing and may be in as many as all four. Put a check mark (√) in the parentheses below a drawing if it contains the *figure*. Put a zero (0) in the parentheses below a drawing if it does not contain the *figure*."

Both procedures were administered in booklet form with separate answer sheets. Testing occurred in individual sessions having a time limit of 10 minutes for each test. The instructions and illustrative examples on the front page of the booklets were read aloud and discussed to ensure that *S* fully understood the nature of the tasks.

RESULTS

During the 10-minute test period provided for Gestalt Completion, *S* undertook as many problems as possible, some responses being correct, some incorrect, and others omitted. In Table 1 are presented the mean frequencies of each type of response for each group. First to be noted is the lack of any appreciable difference between the number of items attempted by the young and old groups. "Attempted" is defined by the last item completed. Thus it appears that both groups at least examined equivalent portions of the total test in the time allotted.

Striking differences are found for the number of drawings correctly perceived. The young have a mean number right of 44.7, in contrast to 15.5 for the old. The t between these values is 7.68 which with 15 df—following Edwards' suggestion of halving the number of degrees of freedom in cases of unequal variance (7, p. 170)—is significant at $p < .001$. Equally sharp differences are seen on number of items incorrectly identified by each group, the mean for the young being 6.0, while that for the old is 15.9. These differences also are highly reliable ($t = 6.11$, $df = 30$, $p < .001$). Moreover, the old group has a significantly greater frequency of failures to respond, 29.4 omissions as compared to 12.4 for young *S*s ($t = 5.44$, $df = 30$, $p < .001$). "Omissions" are simply the number of items left blank of the total number attempted. It is therefore quite clear from these data that the ability of older persons to achieve perceptual closure is distinctly inferior to that of young individuals.

A qualitative analysis of wrong responses offers further insight into the characteristic functioning of the two groups. Almost invariably the types of errors committed by young *S*s consist of inaccurate integrations of the entire picture. That is, the disjointed parts are fused into a whole, but the closure effected is not the intended figure. Sometimes the responses given depart but little from those credited as correct, and quite often the scorer can, as it were, perceive the figure as did *S*. By contrast, older persons usually, though less often, incorporate all the fragments, but in generally loosely organized and sometimes clearly distorted impressions. In addition, although occurring infrequently, all of the old exhibited one or more of the following features in their records: fixation on details of the drawings, perseveration of the same response on successive drawings, and a kind of passive acceptance of the elements as separates, with no genuine closure. In the last category are such answers as "rocks," "clouds," and "flock of birds," which imply a basic deficiency in organizing activity.

Table 2 describes the results for Concealed Figures. It is immediately apparent that the young group completed considerably more items on this test than the old. The difference far exceeds statistical significance ($t = 5.14$, $df = 30$, $p < .001$). However, even within the number of items completed the groups differ significantly in relative adequacy, as measured by the ratio of right minus wrong divided by total attempted ($t = 4.99$, $df = 30$, $p < .001$). This finding is even more impressive in view of the fact that the task seems to be graded in order of increasing difficulty. Thus the greater number of designs attempted by the young necessarily included proportionately more difficult items.

An effort was made to determine whether the types of error provide any further clues as to the differential efficiency of our samples on Concealed Figures. However, the data on hand allow only a very limited approach to this

TABLE 2
MEAN PERFORMANCES OF YOUNG AND OLD GROUPS ON CONCEALED FIGURES

Group	*N*	Attempted	Right	Wrong	Not Perceived When Present	Perceived When Not Present
Young	16	91.1	77.3	13.8	10.3	3.6
Old	16	55.5	36.9	18.6	13.9	4.7

question. A design is scored "wrong" under either of two conditions: 1. the simple figure is not perceived when present in the complex design, and 2. the simple figure is perceived as present in the complex design when such is not the case. For both young and old the first kind of error is decidedly more prominent. Moreover, the findings are similar for the two age groups; the proportion of each variety of error to total errors is approximately equal in both.

DISCUSSION

Perception is conceived to be a resultant of an interplay between field and individual factors. Concerned with demonstrating the importance of closure as one of the major organizing principles in perception, pioneer Gestalt psychologists (12, 13) employed configurations which maximized organizational forces *external* to the person; hence closure was compelling and instantaneous. In an incomplete but "pregnant" field, e.g., a circle with a small arc missing, the patterns themselves uniformly enjoined identical closure. Similarly, with a highly organized and stable field, as in Gottschaldt's original experiments (9), the existing closure was too coercive to segregate parts analytically. In these cases both effecting and resisting closure were viewed primarily as functions of field organization.

However, where field conditions are less coercive the manner and degree of closure is more likely to reflect the organizational skill of the observer. Consider first the problem posed by the Gestalt Completion Test. Here *S*'s task is essentially organizational: given a fragmented drawing he must consolidate the parts and visualize an integrated whole. Variations in the stimuli alone may make integration easier or more difficult, depending on the number, size and location of gaps, and the complexity of the entire drawing; but whether and how closure is achieved depends in large measure on the individual's ability. The situation in the Concealed Figures, which requires *S* to counteract tendencies toward closure, may be viewed similarly. To the extent that factors in the field enhance closures that are in conflict with a particular closure sought, the capacity of the individual to withstand and discard "erroneous" closure is revealed. Thus both the completion of and the resistance to closure are aspects of cognitive functioning which depend on *S*'s general integrative ability (16).

When dealing with Gestalt Completion, not only are the aged shown to be generally less adequate in developing closure, but, more specifically, their responses suggest a loose and/or overly rigid cognitive structure. Percepts are relatively vague and poorly integrated, bearing little resemblance to the figure delineated. They are also focused on details to the exclusion of other aspects of the drawing, perseverative and concretized in the sense that each segment given as a separate is treated as a separate and the relationships between adjoining segments are not envisaged.

Evidences of similar cognitive process in the aged were encountered by us on a task involving the judgment of ambiguous stimuli (14). The *S*s were instructed to respond either "cat" or "dog" to each of a series of 13 line drawings graded through successive small modifications from a clear cat at one extreme to a clear dog at the other. Older persons, in the face of the progressively changing information provided within this order of cards, tended to exhibit a later point of shift in judgment as well as considerably more vacillation in sequence of judgments than younger individuals.

The deficit of aged persons in organizing ability is described perhaps even more strikingly in Rorschach investigations of senescence. Because of the greater ambiguity and lack of structure in the stimuli and the increased demand that organization be provided by *S* it is not surprising that repeated studies have found a distinct constriction and stereotypy of thought, confusion and cautiousness (4, 5, 12, 18).

Unlike the synthetic ability needed on Gestalt Completion, for Concealed Figures *S* must adopt an analytic attitude in order to discover whether the simple figure is embedded in the complex design. Each complex design contains many lesser figures which overlap and are

interwoven in intricate patterns. Identical lines and combinations of lines form one or more sides of a variety of different possible figures having the attribute of good form. The entire configuration itself is a harmonious whole. Thus the task basically taps the ability to maintain and recognize a given closure despite the intrinsic distractions of a more stable larger field and equally good competing gestalten. It is in this particular sense that the aged are found wanting in flexibility of closure.

A special problem arising is that of explaining the matter of omissions. The presence of omissions on Gestalt Completion is probably attributable to the instructions which, it will be recalled, permitted the option of leaving the item blank if it proved too difficult. But our two groups exercised their option in different measure. The old produced a significantly higher frequency of failures to respond than the young. We would hypothesize that this finding reflects one or both of the following characteristics in the aged: 1. greater impotence in the closure function, a phenomenon already witnessed in the analyses of responses actually given; and 2. a more cautious attitude, i.e., the need for a high degree of certainty before committing oneself. This tendency to inhibit response may represent a generalized adaptive mechanism through which the aged person avoids recognizing his inadequacy. Similar behavior in this same group was evidenced on a verbal learning experiment included in our battery (15). In paired-associate learning the older individuals, both initially and on repeated trials, much more usually gave "no response" when the stimulus item appeared.

However, unlike Gestalt Completion, there were no omissions on Concealed Figures. The reason is apparent in the set provided *S* by his assignment. He searches the large design to determine whether it incorporates the smaller figure. If he discovers that it does, he checks the item; if he cannot perceive that it does, then he assumes the figure does not exist within the large design and marks the item accordingly. Under these conditions a failure to respond is therefore not likely.

SUMMARY

A young and old group of 16 *Ss* each, mean ages 26.8 and 78.1 years and equated for intelligence, were compared in performance on two tests of perceptual closure, Gestalt Completion and the Thurstone adaptation of Gottschaldt's Concealed Figures. The former procedure requires the capacity to organize, the latter to resist closure. Older *Ss* were distinctly inferior on both tasks. Moreover, an analysis of "wrong" responses on Gestalt Completion revealed signs of perseveration, fixation on details and concretization by the aged individuals, as well as a more general tendency toward vagueness and poor articulation of form. The old also omitted a significantly greater number of items on Gestalt Completion. This finding was interpreted as indicating one or both of the following: further evidence of a lessened integrative ability, and excessive cautiousness, i.e., a defensive reluctance to venture response for fear of recognizing their inadequacy. The over-all results were viewed as reflecting the overly loose or overly rigid cognitive functioning of the aged.

REFERENCES

1. BABCOCK, HARRIET, & LEVY, L. *Test and manual of directions, the revised examination for the measurement of the efficiency of mental functioning.* Chicago: C. H. Stoelting, 1940.
2. BOTWINICK, J. & BIRREN, J. E. Differential decline in the Wechsler-Bellevue subtests in the senile psychoses. *J. Geront.*, 1951, **6**, 365–368.
3. CALDWELL, BETTYE M. The use of the Rorschach in personality research with the aged. *J. Geront.*, 1954, **9**, 316–323.
4. CHESROW, E. J., WOSIKA, P. H., & REINITZ, A. H. A psychometric evaluation of aged white males. *Geriatrics*, 1949, **4**, 169–177.
5. DAVIDSON, HELEN H., & KRUGLOV, LORRAINE. Personality characteristics of the institutionalized aged. *J. consult. Psychol.*, 1952, **16**, 5–12.
6. DENNIS, W. Age and behavior. *USAF Sch. Aviat. Med. Proj. Rep.*, 1953, Proj. No. 21-0202-0005 (Rep. No. 1).
7. EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
8. GILBERT, JEANNE G. *Understanding old age.* New York: Ronald, 1952.
9. GOTTSCHALDT, K. Über den Einfluss der Erfahrung auf die Wahrnehmung von Figuren. I. *Psychol. Forsch.*, 1926, **8**, 261–317.
10. HOWELL, R. J. Changes in Wechsler subtest scores with age. *J. consult. Psychol.*, 1955, **19**, 47–50.
11. KLOPFER, W. Personality patterns in old age. *Rorschach Res. Exch.*, 1946, **10**, 145–166.
12. KOFFKA, K. *Principles of gestalt psychology.* New York: Harcourt, Brace, 1935.
13. KÖHLER, W. *Gestalt psychology.* New York: Liveright, 1947.
14. KORCHIN, S. J., & BASOWITZ, H. The judgment of

ambiguous stimuli as an index of cognitive functioning in aging. *J. Pers.*, 1956, **25**, 81–95.

15. KORCHIN, S. J., & BASOWITZ, H. Age differences in verbal learning. *J. abnorm. soc. Psychol.*, 1957, **54**, 64–69.
16. KORCHIN, S. J., SINGER, J. L., & BALLARD, R. G. The influence of frustration on the reproduction of visually-perceived forms. *Personality*, 1951, **1**, 54–66.
17. MILES, W. R., & MILES, CATHARINE C. Principal mental changes with normal aging. In E. J. Steiglitz (Ed.), *Geriatric medicine*. Philadelphia: Saunders, 1943. Pp. 99–117.
18. PRADOS, M., & FRIED, E. Personality structure in the older age groups. *J. clin. Psychol.*, 1947, **3**, 113–120.
19. STREET, R. F. *A gestalt completion test*. New York: Teachers Coll., Columbia Univer., 1931.
20. THURSTONE, L. L. *A factorial study of perception*. Chicago: Univer. of Chicago Press, 1944.
21. THURSTONE, L. L. *Mechanical aptitude III. Analysis of group tests*. Chicago: Univer. of Chicago, 1949. (*Psychometr. Lab. Rep.* No. 5.)
22. WECHSLER, D. *The measurement of adult intelligence*. (3rd Ed.) Baltimore: Williams & Wilkins, 1944.

Received March 19, 1956

POSTDECISION EXPOSURE TO RELEVANT INFORMATION[1]

DANUTA EHRLICH, ISAIAH GUTTMAN, PETER SCHÖNBACH,
University of Minnesota
AND JUDSON MILLS
Stanford University

UNDER what conditions will persons voluntarily expose themselves to information? This is an important theoretical and practical question for those interested in social influence processes. It is widely accepted that the audiences of mass media are to a large extent self-selected. However, there has been little rigorous specification or systematic research concerning the variables related to seeking out or avoidance of new information. Festinger has recently proposed a theory (2) which makes some definite predictions about selective exposure to information following decisions, and it is the purpose of this study to test several of them. Before the specific predictions from the theory are stated the relevant parts of Festinger's theory will be summarized briefly.

The theory is concerned with relations among cognitive elements, that is, things which a person knows about himself or his environment. For example, knowing that one is on a picnic is a cognition about one's behavior; knowing that it is raining, a cognition about the environment. Cognitive elements may be either consonant, dissonant, or irrelevant to one another. A behavioral element is consonant with an environmental element if, considering these elements in isolation, the behavioral would follow from the environmental element; dissonant if, considering them alone, it would *not* follow. Thus knowing that it is raining is ordinarily dissonant with going on a picnic; consonant with staying home.

According to the theory, the presence of dissonance gives rise to pressure to reduce or eliminate it. The strength of these pressures is a function of the magnitude of the dissonance. The theory also states that the greater the importance or the value of the elements involved, the greater is the magnitude of the dissonance and hence the greater the pressure to reduce it.

The total magnitude of the dissonance between a behavioral element and the cluster of relevant environmental elements, assuming they are all equally important, is equal to the proportion of dissonant elements in this cluster. Therefore, one way that a person can reduce dissonance is to add new environmental elements that are consonant with his behavior. Thus the theory predicts that persons with dissonance in general seek out consonant information and avoid information that would introduce new dissonant elements.

When a person makes a decision, a corresponding behavioral element is established and his cognitions about the alternatives among which he has chosen are then consonant or dissonant with this element. All the favorable aspects of the unchosen alternatives and unfavorable aspects of the one chosen are dissonant with the choice. Consequently, the creation of dissonance is a common result of decisions. After the choice has been made, the person therefore tends, according to the theory, to expose himself to information that he perceives as likely to support the decision and to avoid information that is likely to favor the unchosen alternatives.

The purchase of a new automobile, for example, is usually a rather important decision for a person. Considerable dissonance should exist for a new car owner immediately after he has bought his car; all "good" features of the makes he considered, but did not buy, and "bad" features of the one he bought are now dissonant with his ownership of the car. He should also attempt to reduce this dissonance. In this instance of postdecision dissonance, sources of information in support of the decision are readily available. Since automobile advertising contains only material favoring the particular car advertised, reading advertisements of his own make is one way a new car owner can get information supporting his

[1] This research was carried out under the direction of Dr. Leon Festinger while the authors were members of his graduate seminar at the University of Minnesota. They would like to thank him for his generous advice and constructive criticism. This study was also part of the research program of the Laboratory for Research in Social Relations.

choice and thereby reduce his dissonance. On the other hand, reading advertisements of other cars may increase his dissonance.

The specific hypotheses of the study were as follows:

1. After a decision persons tend to seek out dissonance-reducing information. Thus new car owners will read advertisements of their own cars more often than those of (*a*) cars which they considered but did not buy and (*b*) other cars not involved in the choice.

2. After a decision persons tend to avoid dissonance-increasing information. Thus new car owners will read advertisements of considered cars less often than of other cars.

3. Postdecision dissonance is, in general, reduced over a period of time. Thus these selective tendencies in readership of advertising will not exist for owners of old cars.

The theory predicts other consequences of postdecision dissonance, for example, changes in desirability ratings of chosen and unchosen alternatives as investigated by Brehm (1). Our present concern is limited to the effect of decision-making on exposure to information.

METHOD

One hundred and twenty-five male residents of Minneapolis and its suburbs, who owned one of eight popular automobile makes, were interviewed[2] in an "advertising survey," after an appointment had been made by telephone. Sixty-five of the respondents were new car owners chosen randomly from a list of recent auto registrations, dated four to six weeks before the time of the interview. The other 60 respondents owned cars manufactured in 1952 (three years before the study was done) or before, and were selected from the Minneapolis telephone directory. (Actually about a third of the old car owners bought their cars after 1952; however, none of the purchases were made during the year in which the study was done.) The number of owners of different makes of cars was approximately equal in both groups.

During the course of the interview:

1. Each respondent was asked to recall recent automobile advertisements which had impressed him particularly and to say where and when he had seen the advertisement, whether or not his impression was favorable, and how much he had read (Recall Data).

2. Each was shown issues of popular magazines which appeared four weeks or less before the interview (after the new car owners had made their decisions) and copies of one of the two Minneapolis newspapers for seven days prior to the interview. Only those publications were included which the respondent previously indicated as having read regularly. For each issue the respondent was asked if he had seen it, and if so, he was shown each car advertisement it contained, asked whether he had noticed the advertisement, and, if he had, whether he had "read all," "read some," or "just glanced at it" (Recognition Data).

3. Each was presented with eight large plain envelopes bearing the names of the eight popular makes. The respondent was told that his reactions to some "new automobile advertising" were desired and asked to choose two of the envelopes in order to read and comment upon the material they contained (Envelope Data).

At the conclusion of the interview the respondent was asked to name other makes which he "seriously considered" before he decided to buy his present car.

RESULTS

As a prerequisite for the predicted selectivity in a person's exposure to information, he must have some expectation about the likelihood that a given item will reduce or increase dissonance for him. Thus the new car owners must first recognize that a particular make is being advertised, i.e., they must at least notice the advertisement. For each respondent the percentage of advertisements which he read, of those which he noticed, was therefore calculated separately for his own car, the cars he considered, other cars not involved in the choice, and also for those considered and other cars combined. (The categories "read all" and "read some" were combined because there were two few advertisements reported as read completely to make a separate analysis feasible; results are essentially the same for these categories.) The means of these percentages for the two groups and the corresponding *N*'s are reported in Table 1, which also contains

TABLE 1

MEAN PERCENTAGES OF ADVERTISEMENTS NOTICED AND OF ADVERTISEMENTS READ OF THOSE NOTICED

(Corresponding *N*'s appear in parentheses)

Car Owners	Make Advertised			
	Own	Considered	Other	Considered Plus Other
	Percentage Noticed			
New	70 (52)	66 (50)	46 (64)	48 (64)
Old	66 (51)	52 (31)	40 (60)	41 (60)
	Percentage Read of Those Noticed			
New	67 (47)	39 (46)	34 (64)	35 (64)
Old	41 (44)	45 (20)	27 (57)	30 (57)

Note.—The *N*'s are reduced because in some cases no advertisements of a particular kind appeared in the issues shown or none of those which appeared were noticed. They are further reduced because not all respondents named cars as "seriously considered."

[2] We would like to thank Craig Erickson and Peter Liebes for their assistance in conducting the interviews.

TABLE 2

SIGN TEST COMPARISONS BETWEEN PERCENTAGE OF ADVERTISEMENTS READ OF DIFFERENT CATEGORIES

Comparison	Car Owners		Comparison	Car Owners	
	New	Old		New	Old
% Own Minus % Considered	+ 21	5	% Own Minus % (Considered Plus Other)	+ 32	20
	− 4	6		− 7	17
	$p < .01$	n.s.		$p < .01$	n.s.
Between Groups $p < .05$ ($\chi^2 = 3.90$)*			Between Groups $p < .01$ ($\chi^2 = 6.89$)		
% Own Minus % Other	+ 31	20	% Considered Minus % Other	+ 19	10
	− 8	16		− 15	7
	$p < .01$	n.s.		n.s.	n.s.
Between Groups $p < .03$ ($\chi^2 = 4.93$)			Between Groups, n.s.		

* Corrected for continuity.

means for the percentage of advertisements noticed of those which were shown.

It can be seen from Table 1 that new car owners read advertisements of their own make more often than advertisements of both considered and other cars. These differences do not appear among old car owners; in fact the difference between reading of own and considered car advertisements is in the opposite direction. However, it is obvious from Table 1 that new car owners do not read advertisements of considered cars less often than of other cars. Finally, it is apparent that readership of advertisements, in general, is greater among new car owners.

The significance of these differences was determined by means of the sign test. For each respondent the sign of the comparison between the percentage of advertisements read of the different categories was found and this distribution of signs was tested using the critical values given in Mosteller and Bush (3). Results of these comparisons appear in Table 2 which shows that the differences apparent from Table 1 are significant. New car owners read significantly more advertisements of their own car than of both considered and other cars; none of the other comparisons yield significant differences. It can be seen that the N's have been further reduced because these comparisons were not possible for all respondents.

The significance of the differences between new car and old car owners was tested by means of chi square. Results of these tests are also given in Table 2. The comparison *own vs. considered* is in the predicted direction with $p < .05$. The comparison *own vs. other* is significant at the .03 level, and the comparison *own vs. considered plus other* is significant at the .01 level; there is no significant difference for the comparison *considered vs. other*.

One other comparison in the recognition data is worth noting.[3] New car owners were divided into two groups; those who named less than two cars as "seriously considered" and those who named two or more. The difference in reading *own vs. considered plus other* was examined separately for these two groups and was found to be greater for those who considered more than two cars. That is, new car owners who named more than one considered car read relatively more advertisements of their own car. For this group the mean difference of the percentages was 37; that of the other group was 26. The difference between the two groups was tested by a median test and the resulting (two-sided) $p = .19$ (12 of the 19 who considered two or more were above the median; only 12 of the 28 who considered less than two were above it).

For the Envelope Data the difference between groups is in the predicted direction but not significant. Fifty-one of 60 (85%) new car owners chose an envelope of their own car, 45 of 59 (76%) old car owners chose one of their own car. Respondents in both groups obviously chose to read and comment upon "new advertising" of their own makes far more often than would be expected by chance.

The number of impressive advertisements (Recall Data) which the respondents read is too small to make analysis of these data possible. (Of course no advertisements were included which appeared before the new car

[3] See the discussion section for the theoretical relevance of this comparison.

owners made their decisions or before a comparable length of time for old car owners.)

DISCUSSION

The data from the recognition of advertisements which appeared in recent publications (after the new car owners made their decisions) strongly support the prediction of Festinger's theory that, after an important decision, persons seek out consonant information. New car owners were found to read advertisements of their own car more often than those of cars they considered but did not buy, and of other cars not involved in the choice. That this selective tendency in exposure to information is not simply a result of car ownership but is related to decision making is demonstrated by its failure to occur among old car owners. The Envelope Data also indicate that new car owners prefer to read advertisements of their own car although the difference between new and old car owners is not significant.

An alternative explanation of the results might be offered. If people do not remember accurately when they read an advertisement, and if shortly *before* buying a car, persons are more likely to read advertisements of *that* car than of others, then the advertisements of their own make may have actually been read before the decision and the obtained differences may be a result of the new car owners' inability to remember the issue in which they read them. This interpretation seems unlikely since it is plausible to suppose that, *before* their purchase, the new car owners attempted to learn as much as possible about all the different makes involved in their choice and therefore read advertisements of the makes that they considered as often as those of the make that they finally chose. Nevertheless, the possibility that it could account for the data was checked by determining whether the advertisements of the make actually chosen by the new car owners which they said they read also appeared during the month prior to their purchases. Only four of 54 such advertisements also appeared in this period. When these few instances in which the alternate explanation might apply are eliminated from the analysis the results are essentially unchanged.[4]

The prediction that new car owners would read fewer advertisements of considered cars than other cars was not substantiated. However, the data are not sufficient to reject conclusively the hypothesis that dissonance-increasing information is avoided after a decision. It is possible that some of the new car owners expected advertisements of considered cars to decrease—not increase—their dissonance. They may have read the advertisements of the rejected cars not with the intention of getting further information about their advantages but for the purpose of finding fault with them, comparing them unfavorably with the chosen car. For example, an owner of a make with high horsepower who regards horsepower as an important feature of automobiles may read advertisements of considered cars of lesser horsepower to remind himself of their inferiority in this respect. We would not expect this to occur frequently but it might happen often enough to account for the failure of the prediction. The data do not provide any test of this interpretation.

The indication that new car owners who named more than one considered car tend to read relatively more advertisements of their own car than do those who named one or none is additional, suggestive evidence for Festinger's theory. Since the total magnitude of dissonance is a function of the proportion of dissonant elements, we would expect that these persons would in general have greater dissonance, for as the number of unchosen alternatives increases the proportion of dissonant elements increases. Therefore the theory predicts that they should show a greater tendency to seek out consonant information.

SUMMARY

Readership of auto advertising by new and old car owners was investigated in order to test some predictions of Festinger's theory of dissonance concerning selective exposure to information following decisions. It was found that new car owners read advertisements of their own car more often than of cars they considered but did not buy and other cars not involved in the choice. These selective tenden-

[4] There are only two small changes in the significance tests reported in Table 2. When the four ads are eliminated the χ^2s for the comparisons between groups for *own minus considered* and for *own minus considered plus other* are reduced to 3.61 and 5.90 respectively.

cies in readership were much less pronounced among old car owners. This finding supports the theoretical derivation that persons in general seek out consonant or supporting information after an important decision in an attempt to reduce dissonance resulting from it.

From the derivation that persons tend to avoid dissonance-increasing information it was predicted that new car owners would read advertisements of considered cars less often than of other cars. The data do not confirm this prediction. It was suggested that some of the new car owners may have expected the considered car advertisements to decrease—not increase—their dissonance.

REFERENCES

1. BREHM, J. W. Postdecision changes in the desirability of alternatives. *J. abnorm. soc. Psychol.*, 1956, **52**, 384–389.
2. FESTINGER, L. The relation between cognition and action. Paper read at Symposium on Cognition, Boulder, Colo., May, 1955.
3. MOSTELLER, F., & BUSH, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. I. Cambridge: Addison Wesley, 1954.

Received March 19, 1956.

A NEW LOOK AT VIGILANCE AND DEFENSE[1]

DONALD P. SPENCE[2]

Teachers College, Columbia University

As New Look experiments continue to accumulate, the exact nature of perceptual vigilance and defense remains open to question. Various investigators (1, 2, 7) have shown that anxiety-provoking stimuli are recognized at both higher and lower thresholds than control stimuli, but the meaning of these threshold deviations is still unclear and their reliability not established. Findings are ambiguous partly because (*a*) the affective qualities of the so-called "taboo" and control stimuli are usually assumed, not predetermined (4, 5, 8); (*b*) the frequency counts generally used often do not apply specifically to the experimental sample; and (*c*) thresholds are usually measured only once, after the experimental threat has been introduced (5, 8). There is thus no check on whether the stimuli were truly matched, and thus no means of partialling out the effect of residual structural differences in "taboo" and control words.

Apart from these procedural shortcomings, the question still remains as to what processes contribute to vigilant and defensive responses. Some studies (1, 2) have succeeded in showing that threshold scores are correlated with personality factors, but the basis for the correlations has not yet been adequately explained. Vigilance and defense have not been explicitly distinguished in terms of the perceptual operations involved, and the possibility exists that they are two manifestations of the same underlying variable.

As a start toward a single-factor theory, it might be argued that both vigilance and defense result from the disrupting effect of the threatening stimuli. Anxiety associated with the "taboo" word, for example, interferes in some way with the *S*'s normal hierarchy of responses to a partially exposed stimulus. This disruption has the effect of reshuffling the usual order of prerecognition hypotheses. As a result, the correct response is given much earlier (low threshold) or much later (high threshold) than it would occur for the matched control word. By this reasoning, anxiety is related to both threshold elevation and depression, and could account for both vigilant and defensive reactions.

These assumptions suggest that words known to create anxiety are recognized at either higher or lower thresholds than control words. The anxiety associated with the threatening stimuli is related to the absolute amount of threshold deviation, but not to either vigilance or defense separately. When the associated anxiety is removed, threshold deviations are expected to diminish.

The present experiment was designed to test these assumptions while taking into account the major sources of error mentioned above. Threat and control words were carefully equated with respect to length, frequency, and ease of recognition before experimental failure was introduced. The experimental anxiety was systematically associated with the threat words, and its effect independently measured. The following hypotheses were tested:

I. Threatening tachistoscopic stimuli are recognized at thresholds *both* higher and lower than thresholds for matched control words. The extent of deviation is related to the degree of anxiety induced by the threat.

II. When the associated anxiety is removed, then threshold deviations diminish. The degree of reduction is related to the decrease in anxiety.

METHOD

Twenty-two adults, five men and seventeen women were tested in individual sessions averaging two and one-half hours apiece. Each session was divided into three approximately equal parts, separated by two ten-minute intermissions. All *S*s were paid for their

[1] This paper is based on a dissertation submitted in candidacy for the degree of Doctor of Philosophy at Teachers College, Columbia University. The writer is indebted to members of his thesis committee, Helen M. Walker, Paul E. Eiserer, and especially E. J. Shoben, Jr., Chairman, for their suggestions and encouragement; to Gardner Murphy who made this study possible by extending his advice, criticism, and the research facilities of the Menninger Foundation; and to David T. Herman and Joseph E. Brewer for making possible initial pilot work at the University of Wichita. This study was done while the writer was a research assistant at the Menninger Foundation.

[2] Now with the Research Center for Mental Health, New York University.

TABLE 1
STIMULUS WORDS

abbot*	hinge	slant*
acorn*	hoist*	snare
baker	kites*	snowy*
beech*	liver	stake
braid	niece*	stung*
click*	orbit*	tarry*
flake*	piper*	tutor
grope*	posed	usher*
handy*	prose	

* Potential failure words which cannot be meaningfully rearranged. Each *S*'s set of failure words was chosen from this group.

time. The majority were persons associated in some way with the Menninger Foundation where the research was conducted—friends and wives of staff; aides; nurses; and outpatients. Five girls in secretarial training completed the group. Whether or not an *S* was an outpatient did not appear to affect subsequent results, so outpatients were included in the final sample.

Twenty-six five-letter words of relatively neutral content and equivalent Thorndike-Lorge frequency were used as stimuli (see Table 1). Sixteen of these words could not be rearranged to form other words (e.g., CLICK) and they constituted the potential failure list. All words were presented initially to the *S* in a random sequence to determine base recognition threshold (Trial I). Words from the failure list (insoluble) were then matched with words of equal threshold from those remaining and all unmatched words were discarded. The matching was done individually for each *S*, providing him with a tailored set of six to eight pairs of words, depending on the number of failure words recognized initially.

Recognition thresholds were determined by gradually increasing the brilliance of a word as it was being exposed at $\frac{1}{100}$ of a second. Change in illumination was controlled by a Variac in parallel with a 75-watt bulb. A Triplett AC volt-meter served as a check on random fluctuations of line voltage. Words were initially exposed with the Variac set at 30 volts and gradually increased two volts per exposure until recognition. Stimuli were typed in capitals and photographed; the resulting black and white negative was fitted into a slide holder and presented 20 inches from the *S*. It was illuminated from behind and mounted in a long oblong box, $28 \times 3\frac{1}{2} \times 3\frac{3}{4}$ inches, open at one end and painted flat black on the inside. Covering the slide was a strip of ground glass to minimize early recognition. Two strips of fluorescent paint outlined the stimulus to orient the *S*. Thresholds were measured in absolute darkness; two five-letter practice words were exposed first to familiarize *S* with the procedure. He was then told to expect 26 additional words of similar length. Each word was presented individually at successive levels of illumination until it was correctly recognized. One correct response was considered sufficient for establishing a threshold.

After all words had been recognized and paired according to their thresholds, each control word was presented to *S* on a white card for 30 seconds. He was instructed to examine the word and mentally rearrange its letters (e.g., DIMLY:MIDLY). He was not directed to form a real word because (*a*) not all control words could be meaningfully rearranged; and (*b*) earlier pilot work indicated that failure to rearrange a word was unsettling, even if *S* were told that it was a practice session. Therefore, *S* was instructed to rearrange each word in whatever way he preferred, as follows:

> This is an exercise in visualizing words. There are no right or wrong answers. It is merely a warm-up session to get you accustomed to thinking of words in different arrangements. The idea is to look at a word and to see in what ways you can rearrange the letters. Obviously, there are a large number of possible combinations you might think of, but I don't expect you to exhaust them all. I merely want you to imagine several arrangements for each word. For example, take the word DIMLY. You might think of it with the D and the M interchanged to make MIDLY. Some arrangements will make sense and others will not; the end result doesn't matter. Take each word as I give it to you and practice rearranging the letters. You don't have to announce each new combination as you find it; make a mental note of the solution and try another. You will have 30 seconds for each word.

The DIMLY:MIDLY example was intentionally presented to form a set which would favor nonsense solutions. This precaution was introduced to discourage *S*s from making meaningful arrangements from the potentially soluble words which might lead to differences in their later recognition. No *S* mentioned a meaningful solution and the chances are they would have if they had found one because the instructions, by implication, suggest there are few. Moreover, no *S* showed signs of boredom or frustration so it may be assumed that the control task was, as intended, essentially neutral.

Another, more difficult, practice task was presented following the control series. Additional words, chosen for their solubility, were given the *S* with instructions to make a new word from each (e.g., SPINE:PINES). After the *S* appeared quite confident of his ability to form anagrams, he was given the failure words with some soluble words added to increase the realism of the task. He was given these instructions:

> All right, your performance on the practice series was excellent. Now I am going to give you a test list. Judging by your performance on the practice list you just finished, you should be able to form one new word from each of the words on the test list. Tell me each word as you find it. You will have 30 seconds for each one.

As the *S* failed each of the failure words, the *E* would sharpen this feeling by asking "Why aren't you doing better? You did well on the practice task," and other, similar questions. Most *S*s became increasingly disturbed by the task and occasionally offered nonsense solutions when they could think of nothing else.

For the majority of *S*s the failure task achieved its intended, depressing effect and its results persevered for the remainder of the experiment (see below). In all other respects the failure and control tasks were essentially similar. In each case *S* was presented with a series of words, each of which he attempted to rearrange within a 30-second period. Words were matched for

length, frequency, and recognition threshold. The *S* was not asked to verbalize his solutions on the control task; he could not do so on the failure task. He tended to form nonsense anagrams on the control task (because of the instructions) and was forced to give similar responses to the failure task (because no other solution was possible). Failure and control experiences were essentially different only with respect to failure, and it seems reasonable to explain subsequent experimental differences in terms of this variable.

Immediately after the failure, each *S* was presented with an Adjective Check List (ACL) of 145 words adopted from Nowlis *et al.* (6). Included in the list were two groups of words identified by Nowlis on an empirical basis: an Anxiety cluster and an Elation cluster of 10 words each. The *S* was asked to check the words which best described his mood. He was then questioned about his feelings, and after a short interpolated task ("List all the words you can think of beginning with Q"), he was asked to list all words he could remember from the failure and control lists.

After a ten-minute rest period during which the experimenter left the room, *S*s' recognition thresholds for the failure and control lists were measured a second time (Trial II). Words were exposed in random order, each to a criterion of one correct recognition. The *S* was then informed that all the anagrams he had failed were actually impossible and that he had been forced to fail because the experimenter was interested in the perceptual consequences of failure. He was encouraged to speak about the experience at some length. When the *S* had expressed his feelings he was given the ACL a second time, and after another interpolated task, memory for failure and control words was again measured. Recognition thresholds were then measured a third time (Trial III) after another ten-minute rest.

RESULTS AND DISCUSSION

Before either hypothesis can be tested, it is necessary to establish that soluble and insoluble control words can be combined for purposes of analysis. Anecdotal reasons for assuming they could be treated together were presented earlier and as a further check, thresholds for soluble and insoluble words were compared on the assumption that if there were differences in the anagram experience they would appear in the recognition scores. Average thresholds after failure for soluble and insoluble control words were compared for each *S*; the difference for the group is not significant ($t = 1.46$; $p > .10$). Meaningful solutions to control words did not seem to inhibit later recognition. Moreover, average failure thresholds were not significantly different from either average thresholds for soluble control words ($t = .68$; $p > .20$) or average thresholds for insoluble control words ($t = 1.58$; $p > .10$). Accordingly, for each *S*, all thresholds for control words were combined and his average control threshold was subtracted from his average threshold for failure words. A second difference score was computed from thresholds for failure and control words after the explanation. A positive score indicates that failure words were recognized, on the average, at higher levels than control words; a negative score indicates the reverse relationship.

Anxiety created by the failure was estimated by counting the number of words checked on the Anxiety cluster of the ACL after failure and dividing by the total number of adjectives checked (Anxiety Ratio I). Anxiety was similarly estimated from the second ACL after the explanation (Anxiety Ratio II). Elation created by the explanation was estimated by counting the number of words checked on the Elation cluster after the explanation and dividing by the total number of words checked.

Memory was scored by subtracting the number of control words from the number of failure words recalled after the failure; a similar score was obtained for recall after the explanation.

In accordance with the first hypothesis, anxiety expressed after the failure (Anxiety Ratio I) was significantly correlated with absolute differences between thresholds for failure and control words after the failure (rho = .57, $p < .01$; see Table 2). Anxiety was uncorrelated with the tendency to recognize failure words

TABLE 2
RANK-ORDER CORRELATIONS BETWEEN MAJOR VARIABLES

	Rho
Anxiety Ratio after failure vs.	
Threshold difference after failure (absolute value)	.57**
Threshold difference after failure (with sign)	.02
Threshold difference after explanation (absolute value)	.52**
Failure word consistency	.38*
Control word consistency	−.22
Memory after failure	.49*
Memory after failure vs.	
Threshold difference after failure (absolute value)	.50**
Threshold difference after failure (with sign)	−.10
Elation Ratio after explanation vs.	
Threshold difference after explanation (absolute value)	.29
Anxiety change vs.	
Threshold change	.11

* $p < .05$.
** $p < .01$.
For 22 cases, a rho of .359 is significant at the 5% level, and a rho of .508 is significant at the 1% level.

above control words (rho = .02); hence, anxiety and "perceptual defense" seem unrelated. On the contrary, *Ss* with high anxiety scores recognized control words at both higher and lower thresholds than failure words, with the absolute degree of difference a function of their anxiety.[5] Table 3 indicates how absolute threshold differences increase as a function of anxiety. Since the failure and control words were equated with respect to frequency, length, and ease of recognition, the differences in perception after failure cannot be attributed to these residual factors. Moreover, suppression of verbal report cannot account for the findings as suggested by Whittaker, Gilchrist, and Fischer (11) because the significant correlation is with recognition deviation, not with recognition delay.

When thresholds alone are considered, recognition of failure and control words seems to be randomly determined. Thirteen *Ss* recognized failure words at average thresholds below control words ("vigilants"); eight *Ss* showed the reverse, or "defense" reaction; and one *S* did not discriminate between lists. For all *Ss*, average failure-word thresholds are not significantly different from average control-word thresholds; $p > .3$; two-tailed sign test (10). The present dichotomy of "vigilant" and "defensive" *Ss* parallels the findings of Postman and Solomon (7), Chodorkoff (2), and others, and supports the general notion that unpleasant stimuli are not necessarily perceived at elevated thresholds. Such data are usually explained by assuming that some *Ss* are "vigilant" and others "defensive." The present findings call this interpretation into question because of the mixed nature of individual responses. No *S* was exclusively "vigilant" or "defensive" in his recognition of all words; rather, each *S* recognized some failure words above, and some below, matched control words.[6] The random distribution of differences for all *Ss* makes it difficult to argue that a "vigilant" or "defensive" average response is characteristic of a particular personality style. The tendency to recognize failure words above or below control words appears to be dependent on factors other than the particular characteristics of the *S*.

TABLE 3
MEAN FAILURE AND EXPLANATION ABSOLUTE THRESHOLD DIFFERENCE SCORES

Group*	Absolute Threshold Difference Score After Failure** (Mean)	Absolute Threshold Difference Score After Explanation*** (Mean)‖
High Anxiety	3.39	3.12
Medium Anxiety	1.68	1.16
Low Anxiety	1.20	.73

* Group placement determined by score on Anxiety Ratio I.
** Between groups, $\chi^2 = 6.6$, $p < .05$.
*** Between groups, $\chi^2 = 9.0$, $p < .02$.
Note.—Significance tests based on an analysis of variance for ranks described in Walker (10).

Absolute size of threshold differences appears to be also important in relation to recall. The tendency to remember more control than failure words after the failure correlated .50 with the absolute threshold difference ($p < .05$). When direction of difference was taken into account, the correlation dropped to −.10. Thus an *S* who forgot relatively more failure words did not necessarily perceive failure words at generally higher levels than control words. Forgetting of failure words was, however, directly related to anxiety. The *Ss* with high anxiety scores tended to recall relatively more control words after the failure (rho = .49; $p < .05$). These results are strongly suggestive of findings by Glixman (3) who tested memory for failure under three degrees of stress. He found that the number of incompleted (failure) tasks recalled varied inversely with the severity of the stress situation; i.e., with the extent to which the instructions stressed the personal implications of the task. Present results bear out his findings to the degree that forgetting of failure (stressful) words is a function of anxiety.

Did threshold differences disappear when anxiety was reduced, as predicted by Hypothesis II? Sixteen *Ss* had smaller anxiety scores on the ACL after the explanation (Anxiety Ratio II) than on the ACL after the failure (Anxiety Ratio I), but only ten of these

[5] It might be argued that more anxious *Ss*, because they seem more sensitive to the differences between failure and control words, would also differentiate between soluble and insoluble control words. They do not. Anxiety after failure is not significantly correlated with the absolute difference between thresholds for soluble and insoluble control words.

[6] The tendency toward recognizing all failure words above or below matched control words correlated −.01 with Anxiety Ratio I. Thus, whether an *S* exhibited a preponderance of "vigilant" or "defensive" responses is not related to anxiety.

had smaller average threshold differences after the explanation. Change in anxiety scores correlated .11 with change in threshold differences. The elation scores correlated only .29 with threshold differences after the explanation. It is possible that the explanation did not produce relief of a sufficient degree to test the hypothesis properly, in which case anxiety created by the failure should still be evident after the explanation. A significant correlation (rho = .52; $p < .01$) between failure-induced anxiety (Anxiety Ratio I) and threshold differences after the explanation suggests that such is the case. Enough anxiety was left unaffected by the explanation to influence thresholds in Trial III (see Table 3). The presence of lingering anxiety is understandable in view of the impact of the failure task and the brief time allotted to explanation.

The high correlation with initial anxiety suggests that because of the cumulative effects of the initial failure, the experiment did not provide a satisfactory test of Hypothesis II. Threshold and memory scores after the explanation are probably some function of both the explanation and the earlier failure, and because of the difficulty in their interpretation, the remainder of the paper will be devoted to an analysis of threshold changes after failure. The high correlation, furthermore, is surprising since one would expect scores on repeated trials to approach some asymptotic limit and argue, in consequence, that since thresholds for control and failure words were originally equated in Trial I, they should again approach equality after repeated testing. That they did not is further evidence of the strength of the experimental failure.

If threshold differences after the failure are related to anxiety, it can be argued that these differences reflect the impact of anxiety upon the perception of the failure words. Accordingly, thresholds for failure words should change more than thresholds for control words from Trial I to Trial II, with change a direct function of anxiety. Further analysis does not support this assumption. Anxiety is not related to change in failure word thresholds, but rather to lack of change, or consistency of recognition. Contrary to expectations, *S*s with high scores on Anxiety Ratio I recognize failure words after failure at thresholds which correlate highly with thresholds before the failure. The higher the anxiety, furthermore, the higher the consistency, and the rank-order correlation between intertrial consistency of failure word thresholds and Anxiety Ratio I is .38 ($p < .05$). Comparable intertrial consistency of control word thresholds, before and after the failure, is not related to Anxiety Ratio I (rho = −.20). Thus it would appear that anxiety is inversely related to failure-word variability and unrelated to control-word variability, suggesting that control-word thresholds can be interpreted as a recognition baseline for all *S*s, unconfounded by the effects of anxiety.[8]

Apparently, then, the anxious *S* after the failure did not substantially alter his choice or sequence of hypotheses in recognition of the failure words. Anxiety stabilized rather than disrupted perception. Although surprising in view of the assumption mentioned earlier that anxiety is disruptive, this finding is in line with recent studies of anxiety and habit strength. Taylor and Spence (9) have suggested that anxiety may reinforce the habit strength of both correct and incorrect responses. Applied to the present study, their findings help explain why *S*s who are anxious following the failure should respond to the failure words essentially as they did initially. Apparently the failure reinforced both correct and incorrect guesses for failure words with the result that recognition thresholds after the failure correlated highly with thresholds before failure. The Taylor and Spence findings might also explain the low intertrial correlation of failure-word thresholds for unanxious *S*s. They presumably responded more flexibly to the failure; some, but not all of their initial responses were reinforced; consequently, their thresholds after the failure were uncorrelated with earlier thresholds.

This line of reasoning needs to be checked by future investigation, but one thing seems clear. The anxiety associated with threatening words apparently leads to an inhibition of

[8] Intertrial consistency of control word thresholds from before to after the failure is not significantly different from intertrial consistency of failure word thresholds over the same period ($p > .3$, two-tailed sign test [10]) which argues against the possibility that recognition of control words after the failure was hindered by associative linkages formed during the control task. Each *S* might have formed more solutions to control words than to failure words during the anagram tasks, but this differential productivity did not seem to affect subsequent recognition.

response; this inhibition may result in either "vigilance" or "defense." What are usually assumed to be two distinctive reactions to tachistoscopic stimuli seem, on closer examination, to be two varieties of the same underlying process. The further fact that no *S* was exclusively "vigilant" or "defensive" in his reaction to the failure words makes it difficult to argue that these ideal response types can be used to discriminate between individuals. If there are distinctive individual ways of reacting to anxiety-provoking stimuli, they seem partly related to the amount of anxiety aroused in the *S*. Further experiments are needed using other measures of anxiety to establish these findings more securely.

SUMMARY

Two lists of words matched for frequency, length, and recognition threshold were presented in an anagram-solving exercise. After one list had been failed, all words were exposed tachistoscopically for a second time. Anxiety aroused by the failure correlated significantly with the absolute difference between recognition thresholds for failure and control words. Anxiety did not correlate with either "vigilant" or "defensive" responses alone, nor were failure words recognized at significantly different levels from control words when the anxiety score was not considered. The relationship between anxiety and absolute threshold difference persisted even after the artificial nature of the failure had been explained, contrary to the hypothesis that explanation would erase the effects of failure.

Sequential analysis of thresholds indicated that successive recognition of failure and control words did not follow a simple pattern. Anxiety appeared to stabilize the repeated recognition of failure words, with threshold consistency a function of anxiety. It can be argued, after Taylor and Spence (9), that anxiety tends to reinforce the use of prerecognition hypotheses for threatening words while not affecting hypotheses for control words. This differential stabilization of failure-word hypotheses seems to underlie both "vigilance" and "defense," suggesting that these response types are less distinctive than is usually assumed.

REFERENCES

1. BRUNER, J. S., & POSTMAN, L. Emotional selectivity in perception and reaction. *J. Pers.*, 1947, **16**, 69–77.
2. CHODORKOFF, B. Self-perception, perceptual defense, and adjustment. *J. abnorm. soc. Psychol.*, 1954, **49**, 508–512.
3. GLIXMAN, A. F. Recall of completed and incompleted activities under varying degrees of stress. *J. exp. Psychol.*, 1949, **39**, 281–295.
4. MCGINNIES, E. M., & ADORNETTO, J. Perceptual defense in normal and schizophrenic observers. *J. abnorm. soc. Psychol.*, 1952, **47**, 833–837.
5. MCGINNIES, E. M., & SHERMAN, H. Generalization of perceptual defense. *J. abnorm. soc. Psychol.*, 1952, **47**, 81–85.
6. NOWLIS, HELEN H., NOWLIS, V., RIESEN, A. H., & WENDT, G. R. Chemical influences on behavior. III. *ONR Tech. Rep.*, 1953, Proj. No. NR 144–060, 19 pp.
7. POSTMAN, L., & SOLOMON, R. L. Perceptual sensitivity to completed and incompleted tasks. *J. Pers.*, 1950, **18**, 347–357.
8. POSTMAN, L., BRONSON, WANDA C., & GROPPER, G. L. Is there a mechanism of perceptual defense? *J. abnorm. soc. Psychol.*, 1953, **48**, 215–224.
9. TAYLOR, JANET A., & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61–64.
10. WALKER, HELEN M., & LEV, J. *Statistical inference.* New York: Holt, 1953.
11. WHITTAKER, EDNA, M., GILCHRIST, J. C., & FISCHER, JEAN W. Perceptual defense or response suppression? *J. abnorm. soc. Psychol.*, 1952, **47**, 732–733.

Received March 19, 1956.

PRACTICE EFFECTS IN REACTION-TIME TASKS IN BRAIN-INJURED PATIENTS[1]

ARTHUR L. BENTON AND HAROLD L. BLACKBURN

State University of Iowa

REACTION-TIME studies of nonpsychotic patients with cerebral injury or disease have disclosed significantly slower times in both simple and choice reactions for such *Ss* when compared with appropriate control *Ss*. However, a considerable degree of overlap in the distributions of scores is typically found, the RTs of at least half of the brain-injured *Ss* being within normal limits (1). The question arises as to whether the presence of cerebral dysfunction in these *Ss* may not reflect itself in a diminished work-capacity, i.e., in a failure to maintain an initial level of performance in respect to speed and stability of response with continued activity. That such is the case has been asserted by Goldstein (2) on the basis of his observations on brain-injured soldiers during and after World War II. He reported that some patients showed initial simple and choice RTs that were well within normal limits. After the first ten trials, however, these patients were likely to show both a lengthening in RT and an increase in variability, whereas normal *Ss* did not show these changes within the limits investigated (30 trials). Thus, a relatively rapid decline in efficiency of performance, as reflected by slower and more variable RTs, appeared to be characteristic of many brain-injured patients whose initial RTs were unremarkable.

The purpose of the present study was to investigate this possibility of differential practice effects in brain-injured and control *Ss* somewhat more systematically than has been done heretofore. Matched groups of patients were run on both simple and choice RT tasks and the findings analyzed to determine what changes with respect to mean score and intra-individual variability occurred in the course of a series of 30 trials on each task.

[1] This investigation was supported by a research grant (B-616) from the National Institute of Neurological Diseases and Blindness, of the National Institutes of Health, Public Health Service. The writers are indebted to Dr. Robert J. Joynt for valuable suggestions and criticisms.

METHOD

Subjects. The *brain-injured* group consisted of 38 patients (28 men, 10 women) on the neurological and neurosurgical services of the University Hospitals, Iowa City, who carried a final and unequivocal diagnosis of supratentorial cerebral disease. A variety of diagnoses was represented in the group, vascular disease and neoplasm being the most frequent (55 per cent). The mean age of the group was 36.9 years (*SD*, 12.0). Mean educational level was 10.0 years (*SD*, 2.6).

The *control* group consisted of 64 patients (52 men, 12 women) on the neurological, neurosurgical, and medical services who showed no evidence or history of cerebral disease, head trauma, or epilepsy. A variety of diagnoses was represented, spinal cord disease being the most frequent (42 per cent). Nine cases were classified as psychoneurotic with somatic complaints. The mean age of the group was 39.2 years (*SD*, 11.5). Mean educational level was 9.8 years (*SD*, 2.3).

No patient who was acutely ill, whose behavior raised the question of psychosis, who had a history of hospitalization for a psychiatric disorder, or whose history suggested mental deficiency dating back to childhood was included in the investigation.

As estimated by *t* and *F* tests, there were no significant differences between the two groups with respect to either the means or variances in age or educational level.

Procedure. The reaction-time apparatus provided for the presentation of a light stimulus which was preceded by a buzz of two seconds duration, the latter stimulus serving as a "ready" signal. Two lights, which were mounted seven inches apart on the right and left sides of an upright panel which faced the patient, were connected respectively to microswitches attached to the right and left sides of the base of the apparatus.

In the simple reaction-time task, the patient, having seated himself comfortably at the table, was instructed to rest the index finger of his right hand lightly on the right microswitch and to press the switch as soon as the right light appeared. The function of the buzz as a "ready" signal was explained to him. Five practice trials were given to acquaint the patient with the task. Thirty test trials then followed.

In the choice reaction-time task, the patient was instructed to rest the index fingers of his right and left hands on the right and left microswitches and to press the corresponding switch when the right or left light appeared. As in the simple reaction-time task, the light stimulus was preceded by a warning signal (buzz) of two seconds' duration The patient was warned not to press both keys at the same time. Lights on the experimenter's side of the apparatus indicated the key which was pressed so that double pressing could be noted by the experimenter. Five practice trials were followed by thirty experimental trials, consisting of

fifteen stimulations on the right side and fifteen stimulations on the left side, presented in a predetermined random order.

The simple reaction-time task was given first and was followed, after a rest period of 2–3 minutes, by the choice reaction-time task. The reaction times were recorded automatically in .01 second units by an electrically driven clock which was activated by the onset of the light stimulus and stopped by the depression of the corresponding microswitch.

RESULTS

Simple RT. The means and standard deviations of the simple RTs in the two groups of *Ss* for successive blocks of 10 trials each are shown in Table 1A and the trends in mean score are graphically represented in Figure 1A. From inspection, it will be noted that while a difference in over-all RT level is obvious, there does not appear to be an impressive difference in the trends over blocks of trials, the control group showing no change and the brain-injured *Ss* showing a slight, questionably significant retardation in RT with continued practice. These impressions were confirmed by the results of an analysis of variance which showed a highly significant main effect for groups ($F = 25.7$ with df 1,100; $p < .001$) but no significant effects either for trials ($F = 0.59$, with df 2,200) or for the groups × trials interaction ($F = 1.11$ with df 2,200).[2] Thus it must be concluded that there is no evidence for significant practice effects, the two trends not being demonstrably different from each other or the combined trend significantly different from a horizontal line.

Three brain-injured *Ss* (8 per cent) showed an increase in RT of 20 per cent or more from the first to the third block of trials. The mean RTs of these *Ss* for the first block of trials were 197, 221, and 273 msec., i.e., two of them showed relatively fast initial RTs. Eight control *Ss* (13 per cent) also showed a retardation in RT of 20 per cent or more from the first to the third block of trials.

Choice RT. The means and standard deviations of the choice RTs in the two groups for successive blocks of 10 trials each are shown in Table 1B and the trends over trials graphically represented in Fig. 1B. It will be seen that while the control group showed a mean retardation of 3 msec. per trial between the first and third blocks of trials, the brain-injured *Ss* showed a mean retardation of 18 msec. However, an analysis of variance indicated that this difference in trend, as estimated by the groups × trials interaction, was not significant ($F = 2.03$, df 2,200). As was the case with simple RT, a highly significant difference in the over-all performance levels of the groups was found ($F = 28.4$ with df 1,100; $p < .001$). The combined trend (trials effect) was not significantly different from a horizontal line ($F = 2.67$ with df 2,200). A separate analysis of variance of the brain-injured trend alone was done to determine if there might be a suggestion of a significant trials effect for this group. The obtained F was 2.29 (df 2, 74), indicating that the hypothesis that the trend represents a horizontal line could not be rejected.

Five brain-injured *Ss* (13 per cent) showed a retardation in RT of 20 per cent or more from the first to the third block of trials, the initial RTs of two of these *Ss* being relatively fast (i.e., under 330 msec.). None of these *Ss* was in the group who had shown noteworthy retardation in RT in the simple reaction task. One control *S* (2 per cent) also showed a retardation in RT of 20 per cent or more from the first to the third block of trials.

Levels analysis of performances of brain-injured Ss. Since clinical observation has em-

TABLE 1

REACTION TIME OF BRAIN-INJURED PATIENTS AND CONTROLS FOR SUCCESSIVE BLOCKS OF TRIALS

(Data in milliseconds)

Statistic	A. Simple RT					
	Brain-injured trials			Control trials		
	1–10	11–20	21–30	1–10	11–20	21–30
Mean	299	304	306	208	204	208
SD	149	141	139	41	45	47
Mean Intra-Individual *SD*	59	65	74	35	33	32

Statistic	B. Choice RT					
	Brain-injured trials			Control trials		
	1–10	11–20	21–30	1–10	11–20	21–30
Mean	371	382	389	303	304	306
SD	97	113	98	55	43	44
Mean Intra-Individual *SD*	71	73	70	55	49	49

[2] The designation, "nonsignificant," is meant to indicate that a result did not attain the .05 confidence level.

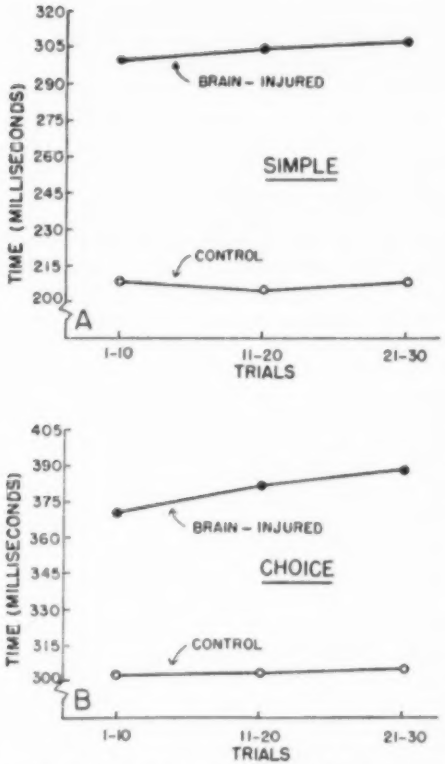


FIG. 1. TRENDS IN MEAN RT SCORES OVER BLOCKS OF TRIALS. A. SIMPLE RT. B. CHOICE RT

phasized susceptibility to fatigue in brain-injured patients whose initial RTs are within normal limits, it seemed indicated to analyze the data further to determine whether such *Ss* might show fatigue effects that were not apparent in the performances of brain-injured *Ss* with initially slow RTs. To make this analysis, the group of brain-injured *Ss* was reduced to an *N* of 36 by random subtraction of two cases and then divided into three subgroups of equal size ($N = 12$), defined by mean RT on Trials 1–10. This was done separately for the simple and choice tasks, with "fast," "medium," and "slow" subgroups in respect to mean RT on Trials 1–10 being formed by the counting-off method. The mean RTs over blocks of trials of these subgroups and the range of RTs for each subgroup on Trials 1–10 for the simple and choice tasks are presented in Table 2. Inspection of this table shows that, apart from an expected slight "regression to the mean" in the extreme groups, there are no important differences in the trends of mean scores over blocks of trials, an impression that is supported by the results of analyses of variance of these trends which failed to show significant levels × trials interaction effects, the *F* for the simple and choice RT interaction being 1.04 and 1.54 respectively, each with *df* 4 and 66.

Changes in intra-individual variability. Since it has been asserted that brain-injured *Ss* tend to become more variable, as well as slower, with continued practice, this aspect of performance was investigated. Intra-individual variability was estimated by computing the mean SDs of the RTs for each block of 10 trials in each of the two tasks for each *S*. The mean SD for each block of trials for the members of each diagnostic group was then computed. These means are shown in Table 1. It will be noted that for both simple and choice RT the brain-injured *Ss* appear to be considerably more variable than the control *Ss*. Moreover, in the simple reaction task, the brain-injured *Ss* appear to show a progressive increase in intra-individual variability while the control *Ss* show no change in this regard. With respect to the choice reaction time no important differences in the trends of the two diagnostic groups are apparent. When these data were subjected to analyses of variance, highly significant group effects were found for both the simple ($F = 13.1$ with *df* 1,100; $p < .001$) and choice tasks

TABLE 2

REACTION TIME OF BRAIN-INJURED PATIENTS FOR SUCCESSIVE BLOCKS OF TRIALS, ACCORDING TO LEVEL OF INITIAL RT

(Data in milliseconds, range is for first block of trials only)

Levels Subgroup	Simple RT		Choice RT	
	Range	Mean	Range	Mean
"Fast"				
Trials 1–10	154–221	195	245–325	297
Trials 11–20		208		292
Trials 21–30		217		310
"Medium"				
Trials 1–10	222–302	256	337–377	358
Trials 11–20		263		385
Trials 21–30		260		396
"Slow"				
Trials 1–10	307–876	466	377–686	477
Trials 11–20		459		480
Trials 21–30		460		475

($F = 17.4$ with df 1,100; $p < .001$). The trials effects (combined trends) were nonsignificant for both the simple ($F = 0.72$ with df 2,200) and the choice task ($F = 0.93$ with df 2,200). With respect to simple reaction time, the groups × trials interaction was significant ($F = 3.77$ with df 2,200; $p < .05$), indicating a difference in the trends of the two groups, the brain-injured *Ss* tending to become more variable than the controls with increasing number of trials. However, no difference in trends was found for the choice reaction task ($F = 0.81$ with df 2,200).

Nine brain-injured *Ss* (24 per cent) and nine control *Ss* (14 per cent) showed an increase of 100 per cent or more in intra-individual variability from the first to the third block of trials in the simple RT task. In the choice RT task, two brain-injured *Ss* (5 per cent) and one control *S* (2 per cent) showed an increase of 100 per cent or more in this characteristic of responsiveness.

DISCUSSION

The control *Ss* show a mean simple visual RT (207 msec.) over 30 trials which seems slightly slower than that which might be expected of healthy *Ss* but is still clearly within normal limits (3). No significant practice effects, either positive or negative, were shown over the three blocks of trials. While the brain-injured *Ss* were characterized by a mean simple visual RT (303 msec.) which is significantly slower than that of the controls, no difference with respect to trend of performance over blocks of trials was shown. Their performance, like that of the control *Ss*, did not appear to reflect significant practice effects. Moreover, this was as true of those brain-injured *Ss* who show fairly fast initial RTs (i.e., over the first block of 10 trials) as it is of those who show a retardation in RT from the very beginning. Thus, insofar as simple visual RT is concerned, no support is found for the expectation engendered by some clinical reports that brain-injured *Ss* may show an undue susceptibility to fatigue.

The findings for choice RT parallel those observed with respect to the simpler task. The brain-injured *Ss* were significantly slower than the controls but did not show a statistically significant differential practice effect. This held for those *Ss* whose initial RTs were relatively fast as well as for the total group. Here, as in the simple RT situation, no substantial evidence for a greater susceptibility to fatigue on the part of brain-injured *Ss* was demonstrable.

The belief that brain-injured patients may become progressively more irregular in RT responsiveness with continued practice is partially supported by the findings. In the simple RT task the brain injured *Ss*, when compared with the controls, showed a significant increase in intra-individual variability over blocks of trials. However, this increase did not occur on the choice reaction task, in which no significant difference in trends was demonstrable.

While the analyzed results of the study are generally negative, it is still true that most of the findings were at least in the direction predicted by clinical report. In the simple RT task, the brain-injured patients showed a mean retardation of 7 msec. from the first to the third block of trials while the control patients showed no retardation. In the choice RT task the brain-injured patients showed a mean retardation of 18 msec. as compared with a mean retardation of 3 msec. in the control group. Moreover, in the simple RT task, a significant increase in variability with continued practice on the part of the brain-injured patients could be demonstrated. These results raise the possibility that modification of the task in the direction of making it more complex, lengthening the foreperiod beyond optimal limits, or increasing the number of trials might yield evidence of differential practice effects. This possibility can be decided only by further investigation.

However, while the study may provide only a limited test of the general hypothesis that brain-injured patients tend to show an undue susceptibility to fatigue in RT tasks, it does constitute a quite adequate test of the validity of the contention, derived from clinical impression, that such patients will show fatigue effects after the first few trials in simple and choice RT tasks. This does not prove to be the case. The findings do indicate that in each diagnostic group there are a few *Ss* who show a noteworthy retardation in RT or a striking increase in intra-individual variability from the first to the third block of trials in either the simple or choice task. It seems likely that clinical observers have been impressed by these changes in responsiveness on the part of some

brain-injured patients but have failed to consider the more numerous negative cases in the brain-injured group or the occurrence of the same type of change in the responsiveness of control patients.

SUMMARY

A study of performance trends in simple and choice reaction time tasks in nonpsychotic, brain-injured patients and a matched group of control patients indicated that, within the limits investigated (30 trials in each task), there were no significant practice effects with respect to mean reaction time for either group in either task. The performances of those brain-injured patients whose initial reaction times were within normal limits also failed to reflect significant practice effects. As compared with the control patients, the brain-injured patients showed a significant increase in intra-individual variability in simple reaction time with continued practice. However, this difference in trends was not found in the choice reaction task.

The findings, while generally nonsignificant, raise the question as to whether a more complex task or a larger number of trials might not yield differential practice effects not disclosed in performance on these simple tasks over a limited number of trials. The results are interpreted as providing no support for the contention of some clinical observers that brain-injured patients are likely to show fatigue effects after the first few trials in simple or choice reaction-time tasks.

REFERENCES

1. BLACKBURN, H. L., & BENTON, A. L. Simple and choice reaction time in cerebral disease. *Conf. Neurol.*, 1955, **15**, 327–336.
2. GOLDSTEIN, K. *After-effects of brain injuries in war.* New York: Grune & Stratton, 1942.
3. TEICHNER, W. H. Recent studies of simple reaction time. *Psychol. Bull.*, 1954, **51**, 128–149.

Received March 21, 1956

A TEST OF THE TAYLOR-SPENCE THEORY OF ANXIETY[1,2]

ELI SALTZ AND ARTHUR J. HOEHN

Air Force Personnel and Training Research Center, Chanute Air Force Base, Illinois

ONE of the most stimulating and provocative developments in recent psychology has been the attempt by Taylor and other Spence students to relate a clinically oriented definition of anxiety to modern behavior theory. Because of the potential significance of this development, it is important that the Taylor-Spence theory be rigorously tested. Unfortunately, some of the early experiments cited in support of the theory were inadequately controlled. The present experiments were designed to evaluate the Taylor-Spence theory under conditions which take into account some of the factors ignored in the early studies.

Taylor (10) originally conceptualized manifest anxiety as a drive variable. Her theory followed the conventional Hullian conceptualization of response strength (R) as a multiplicative function of a learning factor (H) and a drive factor (D). Since anxious *S*s were theorized as operating at a higher drive level than nonanxious *S*s, the prediction was made that more anxious *S*s would learn faster than would less anxious ones. This theory worked out well in simple conditioning experiments; anxiety was found to facilitate the acquisition of conditioned responses (10). The theory proved inadequate, however, for more complex verbal learning. Montague (7) appears to have been the first to use the Taylor scale in a verbal learning study. He reported little evidence to indicate that anxiety facilitates verbal learning. In some conditions, he found that anxiety actually interfered with such learning. As a consequence of the Montague study, Taylor and Spence (11) have recently elaborated the earlier theory.

The revised Taylor-Spence theory holds that in a competing response situation, increased drive should increase the number of responses whose strengths are suprathreshold. Consequently, there should be an increase in the number of erroneous responses which, through the operation of chance fluctuations in response strengths, interfere with the correct response. Since anxious *S*s are assumed, after Miller (6), Mowrer (8), and Taylor (10), to have higher D than nonanxious *S*s, anxious *S*s should perform more poorly than nonanxious *S*s in competing response situations.

Many studies have been performed that seem to support the Taylor-Spence theory (1, 7, 9, 11). However, in all these studies an increase in response competition is accompanied and confounded by an increase in difficulty level. In fact, at times difficulty seems to be used as an *index* of competition, as though no other sources of difficulty exist. For example, in one paper Farber and Spence state, "A further implication of the present theory was that the advantage of the nonanxious over the anxious group would be positively related to the level of difficulty of the individual choice points [of a finger maze], on the assumption that relative difficulty (number of errors) at a choice point reflects the strength of competing incorrect response tendencies at that point" (1, p. 122).

Because of the confounding of difficulty and competition in these studies, any theory based on the hypothesis that anxious *S*s perform more poorly than nonanxious *S*s as a function of the *difficulty* of the task would be upheld by the same data which have been used to date as evidence for the competing response theory.

The present studies attempt to control both competition and difficulty so as to determine their respective effects. In the first experiment, competing and noncompeting verbal learning materials were selected which had empirically been determined to have equal difficulty levels for a group of nonanxious *S*s. The prediction, on the basis of the Taylor-Spence theory, was that the anxious *S*s should do more poorly on the competing material than on the noncompeting, since the increased drive of the anxious *S*s should increase the strength of competing, erroneous responses. The second experiment attempted to compare the performance of anxious *S*s on easy but competing material with their performance on difficult but non-

[1] This research was carried out under the Air Force Personnel and Training Research Center, Chanute Air Force Base, Illinois.

[2] The writers wish to express their appreciation to Professor Lawrence I. O'Kelly for his review of this research and his helpful suggestions concerning it.

competing material. The prediction from the Taylor-Spence theory was that anxious *Ss* should learn faster (relative to nonanxious *Ss*) when competition is reduced, even though difficulty is increased.

EXPERIMENT I

Method

Materials. The use of nonsense syllables permits the independent manipulation of response competition and the difficulty level of material. Both intralist competition and familiarity of the syllables contribute to difficulty in learning a list; greater competition increases difficulty, greater familiarity decreases difficulty. Consequently, it was possible to construct a list of familiar syllables with a great deal of intralist competitiveness which proved to be as difficult for a nonanxious group as a less familiar list with little intralist competitiveness.

The competing list was constructed from Glaze's (2) table of association values. An attempt was made to maximize competition in the following ways: (*a*) The first letter of each word was from the first half of the alphabet. (*b*) The same letter was used as middle letter for each word. (*c*) Three letters, each repeated three times, were used as the first letters of the list.

Lists at various levels of Glaze association value were tried before it was found that, for 10 nonanxious *Ss* who were counterbalanced as to the order of presentation of the two lists, a noncompeting list of 13 per cent association value was as difficult as the competing response list described above. The mean number of trials to a criterion of one perfect recitation was identical through the first decimal place for the two lists. The noncompeting list was constructed so that no two words in the list had the same first letter, and no two words had the same last letter. Letters were from the entire range of the alphabet.

Procedure. Groups of airmen were tested on the Taylor scale, and the upper and lower 20 per cent were selected for use in the experiment. Scores of 22 or above defined the anxious group; scores of 8 or below defined the nonanxious group. These cutting points are almost identical with those used in many of the Taylor-scale studies done at Iowa.

The *Ss* were randomly assigned to learn the two lists of nonsense syllables in the following manner. (*a*) The competing and noncompeting lists were each learned by a different group of 45 anxious *Ss*. (*b*) As a check on the previous equating for difficulty of the competing and noncompeting lists, 18 nonanxious *Ss* learned the competing list and 18 nonanxious *Ss* learned the noncompeting list.

All learning was by the anticipation method on a Missouri-type memory drum to a criterion of one perfect recitation. The syllables were presented at a 2-second rate with 8 seconds between repetitions of a list.

Results

The prediction, based on the Taylor-Spence theory, was that anxious *Ss* should do more poorly on material high in response competition than on material low in response competition if the competing and noncompeting materials were of equal difficulty for nonanxious *Ss*. The results do not sustain this prediction. Table 1 shows the mean learning rates. With 45 anxious *Ss* in each group, $t = .907$ for the difference between competing and noncompeting lists; thus the null hypothesis can not be rejected. In fact, the difference is not even in the direction predicted.

TABLE 1

MEAN NUMBER OF TRIALS TO CRITERION FOR ANXIOUS AND NONANXIOUS SUBJECTS ON "EASY" MATERIAL

Learning Material		Anxious	Nonanxious
Noncompeting 13% association material	Mean	31.09	27.83
	SD	13.59	14.64
	N	45	18
Competing 90% association material	Mean	28.09	28.44
	SD	17.54	12.18
	N	45	18

The original equating of the competing and noncompeting lists for nonanxious *Ss* was cross validated with two new groups of nonanxious *Ss*. The results for these groups, summarized in Table 1, indicate that the initial equating of lists was satisfactory. The nonanxious groups had a difference between means of .61 trials to criterion; $t = .136$ with 34 degrees of freedom.

EXPERIMENT II

Method

Material. For this experiment it was necessary to find a noncompeting list which was more difficult than the competing list used in Experiment I. Since the competing list in Experiment I was as difficult as a noncompeting list at the 13 per cent Glaze association level, the difficult noncompeting list in this experiment was constructed from syllables at the zero association level. Rules for constructing this list were identical with those for constructing the 13 per cent association level noncompeting lists. As was the case in Experiment I, this list consisted of a nine-item series.

Procedure. Eighteen anxious and eighteen nonanxious *Ss* were run on the difficult, noncompeting material. These *Ss* were drawn at random from the same airman population used in Experiment I. Experiments I and II were run substantially simultaneously. The procedure for presenting the list was identical with that described in Experiment I.

Results

The scores obtained by the 45 anxious and 18 nonanxious *Ss* on the competing response list of Experiment I were used as a basis for evaluating the results of Experiment II. The

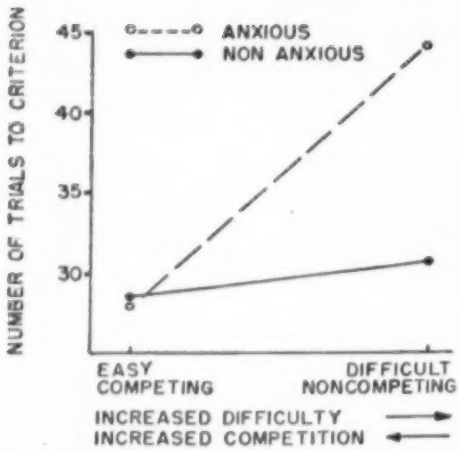


FIG. 1. THE PERFORMANCE OF ANXIOUS AND NONANXIOUS *Ss* WHEN LEVEL OF DIFFICULTY AND DEGREE OF COMPETITION ARE INVERSELY RELATED

prediction from the Taylor-Spence theory was that the anxious *Ss* should do relatively better (compared with the nonanxious *Ss*) on difficult, noncompeting material than on easy, competing material. The results are contrary to this prediction. As can be seen in Fig. 1, the deterioration of performance between the easy but competing and the difficult, noncompeting material was much greater for the anxious than for the nonanxious *Ss*. The difference between these mean differences is 14.50 trials to criterion, significant at beyond the .05 level with $t = 2.05$ and 95 degrees of freedom.

DISCUSSION AND CONCLUSIONS

Neither Experiment I nor II sustains predictions made on the basis of the Taylor-Spence theory. These experiments suggest that the results of previous studies which are often cited as support for the Taylor-Spence theory may be artifacts due to lack of appropriate control over the difficulty levels of competing and noncompeting materials. In Experiment I, the performance of anxious *Ss* was not worse on competing than on noncompeting materials, these materials having been equated on difficulty for nonanxious *Ss*. While the logic of statistics never permits proof of the null hypothesis, the indication that the anxious *Ss* perform no worse on the competing than on the noncompeting material is very convincing. The *N*'s are large (45 anxious *Ss* per group), and the difference is not even in the direction predicted by the Taylor-Spence theory.

In Experiment II, the results are also opposite in direction from those predicted on the basis of the Taylor-Spence theory, and this difference is statistically significant. A problem arises in connection with Experiment II, however. The noncompeting list of this experiment was assumed to be more difficult than the competing list of Experiment I. This assumption was made on the ground that the competing list was as difficult as a noncompeting list of 13 per cent Glaze association level; consequently, a noncompeting list of zero per cent should be more difficult than either the competing list or the noncompeting list at 13 per cent. As can be seen in Fig. 1, both anxious and nonanxious *Ss* do more poorly on the zero per cent noncompeting list than on the competing list, but this difference is not significant for the nonanxious *S*, though it is significant beyond the .01 level for the anxious *Ss*. Even if one assumes, however, that the noncompeting zero per cent list and the competing list are of equal difficulty, the results are contrary to predictions made from the Taylor-Spence theory. From these notions, one would expect the anxious *Ss* to do relatively more poorly on the competing material than on the noncompeting.

These findings give no conclusive idea as to *how* difficulty affects the performance of anxious *Ss*. However, some indication of how difficulty acts is suggested by other recent studies. Lucas (4) reported that the performance of anxious *Ss* deteriorates relative to the performance of nonanxious *Ss* when both groups are told that they are doing poorly on a task. Difficult material may trigger actual anxiety in manifestly anxious *Ss*, and this might somehow produce the decrement in performance. (It should be pointed out that the Taylor scale measures susceptibility toward anxiety, not necessarily the actual presence of anxiety in *Ss*. This point may be important for the interpretation of results obtained with the scale.)

Some recent research relating anxiety and intelligence appears to support the general conception presented here. Mayzner, Sersen, and Tresselt (5), reviewing the literature on the relationship between anxiety and intelligence, found that reported correlations ranged

between +.19 and −.40. They presented evidence which suggests that no correlation between anxiety and intelligence occurs when the intelligence test is taken under nonstress conditions. These writers hold that negative correlations between anxiety and intelligence, such as those reported by Grice (3), may be due to the detrimental effect of stress on anxiety in the test-taking situation.

The above reasoning would suggest that increased difficulty might not be associated with relative decrements in performance by anxious *Ss* if (*a*) *Ss* were reassured of their good performance, or (*b*) *Ss* were in a state of stress while performing on the easy material.

SUMMARY

Previous studies, intended as tests of the Taylor-Spence response-competition theory concerning the effect of anxiety on performance, have confounded amount of competition and the difficulty level of the material. The present studies tested the Taylor-Spence theory under two conditions. (*a*) Competing and noncompeting materials were equated in difficulty for a nonanxious group. The prediction from the Taylor-Spence theory that the anxious *Ss* would perform more poorly on the competing than on the noncompeting material was not sustained. (*b*) Anxious and nonanxious *Ss* were tested on easy competing and difficult noncompeting materials. The prediction from the Taylor-Spence theory was that anxious *Ss* would do more poorly (relative to nonanxious *Ss*) on the easy competing than on the difficult noncompeting material. The results were opposite from those predicted and significant at beyond the .05 level.

The results were interpreted as not supporting the Taylor-Spence theory.

REFERENCES

1. FARBER, I. E., & SPENCE, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 120–125.
2. GLAZE, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–267.
3. GRICE, G. R. Discrimination reaction time as a function of anxiety and intelligence. *J. abnorm. soc. Psychol.*, 1955, **50**, 71–74.
4. LUCAS, J. D. The interactive effects of anxiety, failure, and intra-serial duplication. *Amer. J. Psychol.*, 1952, **65**, 59–66.
5. MAYZNER, M. S., JR., SERSEN, E., & TRESSELT, M. E. The Taylor manifest anxiety scale and intelligence. *J. consult. Psychol.*, 1955, **19**, 401–403.
6. MILLER, N. E. Studies of fear as an acquirable drive: I. Fear as motivation and fear-reduction as reinforcement in the learning of new responses. *J. exp. Psychol.*, 1948, **38**, 89–101.
7. MONTAGUE, E. K. The role of anxiety in serial rote learning. *J. exp. Psychol.*, 1953, **45**, 91–96.
8. MOWRER, O. H. A stimulus-response analysis of anxiety and its role as a reinforcing agent. *Psychol. Rev.*, 1939, **46**, 553–566.
9. RAMOND, C. K. Anxiety and task as determiners of verbal performance. *J. exp. Psychol.*, 1953, **46**, 120–124.
10. TAYLOR, JANET A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.
11. TAYLOR, JANET A., & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **42**, 61–64.

Received April 4, 1956

BREADTH OF DEVIATE CONCEPTS USED BY SCHIZOPHRENICS[1]

LOREN J. CHAPMAN[2] AND JANET A. TAYLOR
University of Chicago *Northwestern University*

In attempting to account for the inferiority of schizophrenics in performing conceptual tasks, one hypothesis that has been advanced, most clearly by Cameron (1, 2), is that schizophrenics are able to form concepts quite readily but are prevented from exhibiting adequate performance by their overresponsiveness to inappropriate stimuli, both internal and external. Two recent studies by one of the present writers have lent support to this hypothesis (3, 4). In these experiments, *Ss* were required to sort cards by common qualities among figures on different cards, there being sufficient communalities between figures that the cards could be sorted in more than one way. In the instructions, *E* designated the "correct" communalities for sorting, thus making the task one of following instructions rather than of discovering the correct concept. The other communalities provided distracting opportunities for sorting in incorrect ways. It was found that schizophrenics, much more than normal *Ss*, tended to ignore the instructions and to sort by these distracting communalities. Further, as the number of these distracters in the test material increased, schizophrenic performance became increasingly inferior to that of normal *Ss* even though the tasks remained essentially the same. This distractibility was found both when the basis of the distracter communality was identity of the figures on the various cards and when the figures possessed communality by virtue of a conceptual category.

These studies showed that in one type of conceptual sorting task, at least part of the schizophrenics' inferior performance arises from overresponsiveness to inappropriate communalities, i.e., communalities other than the instructed ones. The present study investigates interfering overresponsiveness in another kind of conceptual task and seeks to isolate other situational conditions under which it might occur.

It is commonly observed that schizophrenics tend to be "overinclusive" on conceptual sorting tasks, i.e., they include in a given category many instances which are incorrect. In addition, it is often stated that these incorrect instances are frequently related to or superficially resemble objects which properly belong to the category in which they are placed (5, 6). In the present investigation, certain aspects of this latter observation are subjected to experimental test. Specifically, it is hypothesized that the schizophrenic tendency to overinclude in a sorting task is positively related to the degree to which incorrect items share common qualities with the correct items in a given conceptual category.

The method of testing this hypothesis was as follows. Normal and schizophrenic *Ss* were told the name of a common conceptual category (e.g., fruits) and were asked to select from a group of items names of objects belonging to that category. A guide card containing the name of an object in the designated category served as a reminder of the specified category and also of the place where the correct items were to be put. The items from which *Ss* were required to select included, in addition to the correct instances, incorrect instances which were equally divided between members of a category similar to the designated one (e.g. fruits vs. vegetables), and members of a category that was markedly different from the correct one (e.g., birds). It was predicted that schizophrenics would incorrectly include in the designated category more similar items than dissimilar ones. In contrast, normal *Ss*, if they make overinclusive errors at all, should exhibit each type of error equally often.

Should the schizophrenics exhibit such a preference for overincluding similar items, it might be said that their basis for sorting tended to be inappropriately broad, being made according to a wider concept than was specified by the instructions. In the terminology that

[1] The authors are indebted to Drs. Julian H. Pathman and George Faibish of the Downey VA Hospital for assistance in obtaining the schizophrenic *Ss*, and to Drs. Gabriel Ferrazzano and Richard Saavedra of the Chicago USPHS Hospital for assistance in obtaining normal *Ss*.

[2] This study was done while the senior author was a postdoctoral research fellow of the National Institute of Mental Health.

will be used here, schizophrenics, when asked to sort by a given conceptual "species," would be sorting according to a broader "genus," the genus including not only the species specified in the instructions but also a second similar but incorrect species. In the example above, the genus of edibles, which included both fruits and vegetables, would be inappropriately substituted for the designated species of fruits.

In addition to being overinclusive, schizophrenics are also observed to be overexclusive in their sorting, i.e., to exclude from a category instances that properly belong to it. In attempting to account for this type of error, it is possible that such overexclusion results from sorting by a more narrow category than is appropriate. In the present study an attempt was also made to test this suggestion, adding a further hypothesis concerning the conditions under which such narrowing might occur. A summary of the actual experimental procedure might best serve to introduce these hypotheses. Using the same task materials as in the previously discussed sortings, *E* told *S* the name of a conceptual category that was, in the terminology used above, a "genus" and asked *S* to select from a group of items names of objects belonging to that category. The correct items available for sorting included equal numbers of members of two constituent species categories. The incorrect items were, as before, members of a category that was dissimilar from the genus or its two species categories. A guide card with the name of an object from the genus category was used to indicate the place for the genus cards. Naturally, the object on the guide card could belong to only one of the two species groups. It was expected that if errors of exclusion in schizophrenics are due to excessive narrowing of the conceptual basis of sorting, such narrowing would be most likely to occur in the situation in which a broad category, reducible to smaller ones, was employed. Further, in line with the notion that schizophrenics are highly distracted by external stimuli present in the situation, the direction of narrowing might be toward sorting according to the species to which the guide card belonged. Thus, schizophrenics were expected to exclude incorrectly from the genus sorting fewer items in the species category to which the guide card belonged than items in the species category to which it did not. Again it was anticipated that such errors as normals might make would reveal no differential tendency toward the two types of overexclusion.

In one of the previously discussed studies of distractibility by Chapman (3), the criteria for choosing *S*s were such that only moderately impaired schizophrenics could be used. Nonetheless, when the group was divided into two groups according to severity of illness, significant differences were found, the milder cases exhibiting fewer errors. The difficulty level of the tasks in the present study was such that more severe cases could be accepted. In order to determine whether degree of illness could again be related to the amount of performance impairment, two groups of schizophrenic *S*s were tested, one judged to be relatively severe in degree of impairment and the other, comparable to the total group previously studied, moderately impaired.

METHOD

Subjects. One normal and two schizophrenic groups, each consisting of 24 males, were tested. The normal *S*s were drawn from the general medical and surgical patients at the USPHS Hospital, Chicago, Illinois. No individual was included in this group who appeared by hospital record or by interview to have a psychiatric or neurological disorder. The schizophrenic groups, patients in the Downey VA Hospital, had all received a hospital diagnosis of chronic schizophrenia, with their first hospitalization for psychosis occurring at least one year previously and their last at least 3 months previously. Other prerequisites for selection were absence of any indication of cerebral organic involvement or mental deficiency, no current drug therapy, no EST within the past 3 months, or more than 35 EST treatments altogether. The two schizophrenic groups were distinguished by their ward status, the "moderately incapacitated" group being drawn from patients with ground privileges or assignment to an outside work detail, and the "severely incapacitated" group being drawn from patients with neither ground privileges nor outside work assignments. In addition to the 24 *S*s in each of the schizophrenic groups, 4 moderate and 18 severe schizophrenic *S*s were also seen but were not included in the final groups because they were unwilling or unable to participate in the experimental procedure.

Mean age was 37 years for each of the three groups, while mean years of education were 10.4 for the moderate schizophrenics, 10.6 for the severe and 9.8 for the normals. Although the groups did not differ significantly with respect to these two variables, both schizophrenic groups were significantly lower on functioning IQ as determined by the CVS, presumably because of the disruptive effects of schizophrenic pathology. On this test, the mean IQ was 107 for the normals, 88 for the moderate schizophrenics, and 92 for the severe.

TABLE 1

THE CATEGORIES AND GUIDE CARD USED IN EACH TASK, AND THE NAMES USED IN THE INSTRUCTIONS FOR THE CONCEPTUAL SPECIES AND GENUS

Task	Categories	Guide Card	Name for Species Category	Name for Genus Category
A	kitchen utensils furniture birds	pan	kitchen utensils, i.e., utensils used for cooking or eating	things found in a house
B	alcoholic beverages nonalcoholic beverages insects	wine	alcoholic drinks, i.e., drinks that if you drink enough of they make you tight	things that a person can drink
C	trees flowers musical instruments	oak	names of trees	growing things
D	fruit vegetables tools	apple	fruit	things that a person can eat
E	domesticated animals wild animals recreational equipment	cow	domesticated animals, i.e., animals that are kept on the farm or in the home	animals
F	wearing apparel household linens sea creatures	skirt	wearing apparel	things made of cloth

Materials. Six sorting tasks were devised, each consisting of 30 words typed on separate 3″ × 5″ index cards. The words were equally divided among three conceptual categories, two of which were designed to be similar but not overlapping (e.g., alcoholic and nonalcoholic beverages) and the third markedly different from the other two (e.g., insects). Each of the two similar categories is here called a "species" while the two together are called a "genus." The third, different, set of words is considered as control items. Before the experiment proper, the materials were given to a number of normal *Ss*, similar in background to those used in the experiment, in order to insure that for such *Ss* the items used were known and could readily be assigned to the proper category. The actual categories used in each of the tasks are listed in Table 1.

A 13″ × 10″ × 4″ box was used for sorting the cards, the box having two slots in the top through which the cards could be deposited. A wire rack for holding a guide card was placed on top of the box in back of one of the slots.

Procedure. Each *S* was required to sort three of the six tasks according to a conceptual species, i.e., to separate the designated species cards from the second species and the dissimilar control items. The remaining three tasks were sorted according to a conceptual genus, i.e., separation of the genus category, made up of the two species, from the control items. Each of the six tasks could be used for either type of sorting, with only the instructions distinguishing the two.

For both types of sorting, *E* handed *S* the 30 cards of that task, named the category of cards that were to be separated out from the others and placed a guide card containing a word illustrative of the specified category on the rack in back of the slot in which cards of that category were to be placed. Instructions given on each task may be illustrated by the following, taken from the species directions for Task D (See Table 1):

> This card up here says "apple." An apple is a fruit. Take these cards and each one that names a fruit goes in this slot beneath the apple. Everything else, that is, everything that is not a fruit goes in this slot here.

Except for the name of the particular category and the guide card used, these same instructions were used for all tasks on both types of sorting. The specific designations and the guide cards employed for both species and genus sortings are listed in Table 1.

Each of the six sorting tasks was used for species sorting with half of the *Ss* in each group and for genus sorting with the remaining. The order of presentation of the two types of sortings was of the abbaab pattern for each *S*, half of the *Ss* beginning with the species sort. The position of the guide card on the box was also varied for each *S*, alternating from right to left on successive tasks. Finally, the particular order of the six tasks was systematically randomized among *Ss*.

RESULTS

Species sorting. The distributions of all types of errors were extremely skewed, especially for the schizophrenic groups. For this reason, both mean and median error scores were computed for each group and all statistical tests were of the nonparametric type.

As predicted, the schizophrenics incorrectly placed in the designated species category more of the similar species items than dissimilar control words, while the normals made few errors of either type. (More specifically, as indicated by the median error scores shown in Table 2, the severe schizophrenics made 6 out of a possible 30 incorrect placements of the similar species items, the moderate group 6, and the normal *Ss* a median of zero errors. When compared by means of a Mann-Whitney

test, highly significant differences were found between schizophrenic and normal *Ss*, the comparison between the normal and moderate groups yielding a z of 2.76 ($p < .01$) and the comparison of normals and the severe schizophrenics a z of 2.78 ($p < .01$). No significant difference was found, however, between the two schizophrenic groups ($p = .43$). In contrast to these normal-schizophrenic differences for the similar items, the median number of errors of placement of the dissimilar cards was zero for all three groups and no significant differences were found among the groups.

One additional comparison was made in order to obtain a more direct measure of the tendency of the schizophrenics to overinclude differentially on the basis of similarity. For each *S*, the number of incorrectly placed dissimilar items was subtracted from the number of incorrectly placed similar items. As may be seen in Table 2, the median value of this difference score is zero for the normal group but much higher for the schizophrenic groups. (Examination of the data reveals that the lower difference score for the severe schizophrenics, as compared to the moderate group, is largely due to the tendency of the severe *Ss* to make more errors of *both* types, thus lowering the magnitude of the difference between the two types of errors.) Comparisons of the groups on these difference scores by means of a Mann-Whitney test yielded a z value of 4.10 ($p < .001$) between normals and moderate schizophrenics and of 3.44 ($p < .001$) between normals and the severe group. Again, the two schizophrenic groups did not differ significantly ($p = .74$), both showing an equal tendency to include incorrectly similar items. Inspection of the errors on each of the six tasks separately revealed that this differential error tendency of the schizophrenics occurred on each of the tasks.

TABLE 2
MEDIAN AND MEAN NUMBER OF ERRORS ON THE SPECIES SORTINGS

Group		Inclusion			Exclusion
		Similar	Dissimilar	Diff. (S-D)	
Normal	Mdn.	0	0	0	0
	M.	.54	.33	.21	.58
Schizophrenic; moderate	Mdn.	6.00	0	6.00	1.50
	M.	7.33	.29	7.04	3.04
Schizophrenic; severe	Mdn.	6.00	0	3.50	1.00
	M.	8.20	1.33	6.87	2.67

The schizophrenic groups differed from the normal group, it will be remembered, with respect to IQ scores, even though they were matched on education. Therefore the observed tendency of the schizophrenics to overinclude items differentially on the basis of similarity might be attributed to a lowering of functioning intelligence. In order to evaluate this possibility, 19 normal and 24 schizophrenic *Ss* were selected from the total groups in such a way that the mean and *SD* of the IQ scores for the two groups did not differ significantly. The mean IQ of these selected normal *Ss* was 105, and of the schizophrenics, was 104. When these groups were compared, the median score for the difference between the dissimilar and similar types of error was 0 for the normal and 2 for the schizophrenic *Ss*, while the means for the two groups were .1 and 4.8 respectively. The difference between the two groups, tested by a Mann-Whitney test, was significant at the .001 level of confidence, thus indicating that the tendency to overinclude is not accounted for by differences in functioning intelligence.

As a matter of incidental interest, attention might be called to a different type of error than those mentioned above: the number of correct species cards *excluded* from the designated category. From Table 2 it may be seen that the discrepancy between the normal group and each of the schizophrenic groups on this score was not as striking as on the overinclusion measures, but in both cases the difference was significant at the .01 level.

Genus sorting. In the genus sorting, it will be recalled, correct sorting consisted of placing with the guide card all the members of a conceptual genus which consisted of two conceptual species. It was expected that schizophrenics might exhibit a differential tendency to *exclude* items from the designated genus category, i.e., to omit fewer of the items falling into the same species category as the guide card than items belonging to the second correct species. Contrary to expectation, however, the frequency of errors of exclusion was not related to the similarity of the items to the guide card. Although, as seen in Table 3, both schizophrenic groups excluded a few less of the species category to which the guide card belonged, the difference in frequency of the two

TABLE 3
MEDIAN AND MEAN NUMBER OF ERRORS OF INCLUSION AND EXCLUSION OF GUIDE-CARD AND NON-GUIDE-CARD SUBCATEGORIES FOR GENUS SORTINGS

Group		Exclusion			Inclusion
		Guide-Card Cat.	Non-Guide-Card Cat.	Diff. (NG-G)	
Normal	Mdn.	0	0	0	0
	M	.25	.25	.00	.21
Schizophrenic; moderate	Mdn.	0	0	0	0
	M	.83	1.12	.29	1.42
Schizophrenic; severe	Mdn.	.50	1.00	0	0
	M	1.96	2.50	.54	2.00

types of error, examined by a Wilcoxon test, did not approach significance for any of the groups.

Errors of any type were, in fact, relatively infrequent in all of the groups on this genus sorting, most of the medians being zero. Further, when the groups were compared on each of the two types of errors of exclusion and on the errors of incorrectly including items from the third, incorrect category, none of the groups differed significantly. These figures, however, seem to overestimate slightly the goodness of the schizophrenics' performance. When groups were compared on total errors of exclusion, the severe schizophrenic group was found, by the Mann-Whitney test, to exceed the normal group significantly ($p = .02$). No other group comparison on this measure was significant, however. This general tendency toward exclusion that appeared in at least the severely incapacitated schizophrenics was paralleled, it will be recalled, by a similar tendency in the species sorting in which both schizophrenic groups significantly exceeded the normal *Ss* in their errors of exclusion. Thus, differences were found between normal and schizophrenic groups on overexclusion measures although the basis for this behavior has not been isolated.

The slightness of the differences between groups on the genus sorting, as compared with the differences on the species sorting, suggested that the genus task was less difficult for the schizophrenics than the species task. In order to check this more directly, the total number of errors made in each group on the two types of tasks were compared by a Wilcoxon test. Both schizophrenic groups were found to make more errors on species sorting than on genus sorting. For the moderately incapacitated, the z value was 4.04 ($p < .001$) and for the severe, the z was 2.83 ($p < .01$). For the normal *Ss*, however, there was not a significant difference ($p = .73$).

DISCUSSION

The analysis of errors of inclusion in the species sortings confirmed the hypothesis that schizophrenics would incorrectly include more items from the category similar to the correct one than from the dissimilar category. In attempting to describe the overinclusive behavior of these *Ss*, it might be said that the schizophrenics continued to perform conceptually but according to a broader category than was specified by the instructions. These results thus add further support to the contention that schizophrenic inferiority on conceptual tasks is due to overresponsiveness to inappropriate stimuli rather than loss of conceptual ability per se.

The hypothesis that on the genus sortings schizophrenics would show the opposite reaction—excessive narrowing of the conceptual basis for sorting—received no support. Schizophrenics made more errors of exclusion than did normals (although only the severe group differed significantly from normal *Ss*), but the attempt to isolate stimulus conditions affecting these errors was unsuccessful.

One of the interesting results of the genus sorting in general was the fact that schizophrenics made so few errors of any type, inclusive or exclusive. In a sense the comparative adequacy of the schizophrenics' performance further confirms the findings of the species sorting: when faced with separating categories of items that are dissimilar (as in the genus tasks), schizophrenics differ comparatively little from normals; when the correct items are relatively similar to some of the incorrect, schizophrenic performance suffers although that of normals does not. Related to this is the unexpected finding that schizophrenics showed a more pronounced tendency to exclude correct items from the designated category in the species than in the genus sortings. A possible reason for this is that in the presence of stimuli that act strongly as distracters, schizophrenics become more susceptible to distraction by all sorts of other stimuli, internal and external, which under other circumstances do not have much influence.

Finally, a comparison can be made of the

two schizophrenic groups. The severely incapacitated group tended to make more errors of all types on both kinds of tasks than the moderately incapacitated and included more *Ss* with high error scores, as can be seen especially from the mean error values shown in Tables 2 and 3. Although the difference between groups in no case reached an acceptable significance level, there was, at least, a consistent direction of difference, the severely incapacitated *Ss* tending toward a more inferior performance.

SUMMARY

The present study attempted to isolate stimulus conditions that result in certain types of errors being made by schizophrenics on a conceptual sorting task. Normal and schizophrenic groups were asked to sort words naming objects falling into three conceptual categories, two of which were similar to one another, and the third dissimilar. In the first part of the study *Ss* were asked to separate items belonging to one of the similar categories from the remaining items. It was expected that for the schizophrenics, but not the normal *Ss*, the items incorrectly included would more often be from the similar incorrect category than from the incorrect dissimilar one. The results confirmed this expectation, schizophrenics differing significantly from normal *Ss* on a score of number of incorrect similar items included minus number of incorrect dissimilar items. These errors of the schizophrenics therefore may be described as resulting from an inappropriate broadening of the conceptual basis of sorting.

An attempt was also made, in the second part of the study, to determine whether an opposite kind of error, excessive narrowing of the conceptual basis for sorting, might also take place under certain circumstances. The groups were again given items falling into three categories but the designated category which they were asked to separate from the other items consisted of the two similar subcategories. It was expected that schizophrenics, unlike normal *Ss*, might tend to sort according to only one of these subcategories, thus excluding incorrectly more of one than the other. This prediction was not confirmed, schizophrenics showing little difference in performance from normal *Ss* in these sortings.

A comparison was also made of the performance of severely and moderately incapacitated schizophrenics. Although the severe group was in no instance significantly different from the moderately incapacitated, the direction of difference between the groups pointed quite consistently to the relative inferiority of the severe group.

REFERENCES

1. CAMERON, N. Reasoning, regression and communication in schizophrenia. *Psychol. Monogr.*, 1938, **50**, No. 1 (Whole No. 221).
2. CAMERON, N. Deterioration and regression in schizophrenic thinking. *J. abnorm. soc. Psychol.*, 1939, **34**, 265–270.
3. CHAPMAN, L. J. Distractibility in the conceptual performance of schizophrenics. *J. abnorm. soc. Psychol.*, 1956, **53**, 286–291.
4. CHAPMAN, L. J. The role of type of distracter in the conceptual performance of schizophrenics. *J. Pers.*, in press.
5. GOLDSTEIN, K., & SCHEERER, M. Abstract and concrete behavior: an experimental study with special tests. *Psychol. Monogr.*, 1941, **53**, No. 2 (Whole No. 239).
6. RAPAPORT, D., GILL, M., & SCHAFER, R. *Diagnostic psychological testing*. Vol. 1. Chicago: Year Book Publishers, 1945.

Received April 13, 1956.

CRITIQUE AND NOTES

EFFECT OF PREINFORMATION UPON HUMAN CHARACTERISTICS ATTRIBUTED TO ANIMATED GEOMETRIC FIGURES[1]

RONALD E. SHOR

Brandeis University

THE present study attempts to demonstrate that the impression that a person forms of another individual's personality is partially determined by the characteristics of the other people with whom he is seen to interact. Following the work of Asch (1) and others (2, 3, 5, 6, 7, 8, 9), who have demonstrated that preinformation about an individual's personality affects first impressions of him, this study attempts to show that such preinformation also affects impressions of the personalities of others who interact with him. Animated geometrical figures were used instead of actual persons.

In the Heider-Simmel movie (4) which provided the stimulus material, three geometric figures (a large triangle, *T*, a small triangle, *t*, and a small circle, *c*) move about and interact in such a way as to lead most subjects (*Ss*) to report the action as that of three "persons" acting out a story. Most *Ss* spontaneously perceive *T* as an aggressive, "nasty" male who picks a fight with *t* and who attempts to take advantage of *c*; *t* is seen as a spirited, heroic male who attempts to protect *c* from *T*; *c* is seen as a fearful, defenseless female who seeks the protection of *t*.

It was hypothesized that if *Ss* were given different preinformation about the personality of one of the figures (*T* was chosen) the way in which the other two figures are perceived would also be affected. Since *Ss* tend spontaneously to perceive *T* unfavorably it was hypothesized that giving them unfavorable preinformation about *T* would either have no additional effect or would tend to accentuate slightly the way in which they spontaneously perceive the three figures. That is, if there were any changes, *T* would be seen as "nastier," *t* as braver, and *c* as more defenseless. On the other hand, it was hypothesized that favorable preinformation about *T* would cause a considerable change in interpretation. The *Ss* would perceive *T* more favorably and perceive *t* and *c* less favorably.

METHOD

In a pilot study 34 members of an undergraduate psychology course at the University of Kansas were shown the Heider-Simmel movie twice and asked to describe the personalities of the three figures. Before seeing the movie, however, half of the *Ss* were told that *T* was an aggressive person disliked by most people, and the other half were told that *T* was a fair minded person liked by most people. From the *Ss*' reports the investigator listed all of the trait-names used to describe one or more of the three figures. This list was then reduced to 40 trait-names on the basis of frequency of report and diversity of coverage.

In the present experiment, 300 students from an introductory psychology course were shown the Heider-Simmel movie twice and then asked to check for each of the figures those of the 40 trait-names which were felt to be relevant. The conditions were identical for all *Ss* with the exception that 100 *Ss* received the aggressive-disliked preinformation about *T*'s personality (AD group), 100 others received the fair-minded and liked preinformation (FL group), and 100 more received no preinformation, for purposes of control (Cn group).

The instructions for the AD group were as follows: "We are going to show you a movie in which there are three geometric forms; a large triangle, a small triangle, and a circle, moving in various ways in and around a larger square. Please imagine that the two triangles and the circle are people acting out a story. We will show you the movie twice and then ask you to write out the answers to a number of very simple questions about the movie. In no way is this a test of your intelligence or personality, and there is no time limit. The only bit of information about the movie which we are giving you is that the large triangle is an aggressive person who is disliked by most people. Remember this fact and use it to help understand and interpret the story: *the large triangle is an aggressive person who is disliked by most people.*"

Instructions for the FL group were identical except that the words fair-minded and liked were substituted for the words aggressive and disliked. In the Cn group the last two paragraphs were deleted.

RESULTS

In Table 1, the trait-names are arranged according to the size and direction of difference between the AD and FL groups in the frequency of their choice as descriptions for *T*. Since the size of each group is 100, these frequencies are also percentages. To illustrate, 92 *Ss* in the AD group checked the trait-name "not likable" for *T* in comparison to only 24 *Ss* in the FL group. The difference of 68 between these two frequencies is the largest in which an AD frequency exceeded an FL frequency, and thus "not likable" is ranked in

[1] This report is based on a dissertation submitted in May, 1955, in partial fulfillment of the requirements for the M.A. degree at the University of Kansas, under Dr. Fritz Heider, thesis advisor.

TABLE 1
FREQUENCIES FOR EACH FIGURE OF TRAIT ASCRIPTION IN THE AD AND FL GROUPS

Trait-names	Figure											
	T			*t*			*c*		
	AD	FL		AD	FL		AD	FL	
Not likable	92	24	*	1	23	*	2	9	
Troublemaker	85	29	*	13	52	*	2	12	*
Bully	91	45	*	2	13	*	0	0	
Bad person	57	15	*	1	18	*	0	6	
Showoff	68	29	*	13	30	*	2	5	
Treacherous	43	11	*	0	12	*	0	6	
Defiant	48	22	*	21	31		5	8	
Boastful	43	21	*	4	16	*	0	2	
Unsure of self	48	26	*	18	31	*	55	52	
Aggressive	90	70	*	23	35		4	5	
Conceited	36	16	*	3	12	*	0	6	
Cowardly	25	8	*	12	19		29	28	
A tease	26	14		28	33		13	19	
Hot-tempered	97	86	*	2	19	*	0	6	
Tricky	23	12		20	38	*	2	8	
Fickle	14	8		2	9		22	27	
Curious	15	14		32	17	*	40	31	
Defenseless	2	1		21	24		89	70	*
Timid	0	0		17	17		82	76	
Dependent	13	14		15	17		63	65	
Male	77	79		67	74		5	11	
Persistent	66	69		44	44		4	4	
Gullible	16	21		4	8		18	28	
Female	4	10		9	8		76	72	
Fun-loving	1	8		32	25		33	23	
Minds own business	2	9		30	7	*	48	27	*
Peaceable	0	8		45	17	*	67	38	*
Clever	8	16		41	38		17	10	
Playful	4	13	*	45	28	*	31	22	
Honest	1	10	*	35	7	*	25	13	*
Loyal	0	10	*	59	28	*	31	29	
Good natured	0	13	*	44	17	*	39	21	*
Likable	0	14	*	63	23	*	62	40	*
Courageous	5	23	*	59	31	*	8	3	
Brave	4	24	*	61	33	*	6	3	
Fair-minded	0	21	*	35	11	*	27	14	*
Good person	1	24	*	66	30	*	65	33	*
Has spirit	40	68	*	78	72		27	24	
Protective	5	33	*	66	36	*	8	7	
Easygoing except when provoked	14	58	*	43	17	*	11	6	

* Difference significant at $p < .05$.

TABLE 2
AVERAGE FREQUENCIES OF FAVORABLE AND UNFAVORABLE TRAIT ASCRIPTION FOR EACH FIGURE

Experimental Group	Figure					
	T		*t*		*c*	
	Fav.	Unfav.	Fav.	Unfav.	Fav.	Unfav.
AD	6.2	63.5	54.5	8.7	28.3	7.6
FL	25.9	30.9	27.8	23.9	17.9	11.1
Cn	8.1	55.5	49.8	12.2	28.2	7.9

first place. An asterisk indicates that this difference is significant.[2] The trait-name ranked last is the one in which the greatest excess of an FL over an AD frequency occurred.[3]

As can be seen from the table, where more *Ss* in the AD group than in the FL group see *T* as possessing a certain trait, a very strong tendency exists for fewer AD than FL *Ss* to see this trait in the other two figures. The inverse is also true. Moreover, differences between AD and FL group frequencies are greatest for *T*, fairly large for *t*, and small but observable for *c*. Where the differences tend to be large for *T* they also tend to be large for *t* and *c*, and where they are smaller for *T* they tend to be smaller for the others also. Thus there is a clear and consistent tendency for impressions of *t* and *c* to vary together and in an opposite fashion from impressions of *T*.

The trait-names were dichotomized by the investigator according to whether they indicated favorable or unfavorable impressions of personality. In Table 2, frequencies for the unfavorable and favorable trait-names in which significant differences between the AD and FL Gruops were found with respect to *T* are combined to yield a favorable and unfavorable average frequency for each figure for the three experimental groups.

In the Cn group where the stimulus configurations of the movie operate alone without any preinformation, *T* was viewed in a generally unfavorable light; *t* and *c* in a generally favorable light. These findings conform to expectations. In the AD group provided with unfavorable preinformation about *T* these same relations were preserved, and, except for *c*, were intensified. In the FL group, however, where the preinformation about *T* is contrary to the way most *Ss* spontaneously perceive the film, *T* was still more frequently seen as unpleasant than pleasant but many more *Ss* perceived *T* as possessing favorable traits and fewer perceived him as possessing unfavorable traits than in the AD group. The other two actors were also still seen as more pleasant than unpleasant

[2] Significance was accepted at the 5 per cent level. In most cases, however, probabilities were considerably less than the 1 per cent level.

[3] Control (Cn) group frequencies are not included in this table. In most instances Cn frequencies are in between AD and FL frequencies, although much closer to the AD frequencies. To save printing costs, a table giving all of the data and the results of tests of significance for all three experimental groups has been deposited with the American Documentation Institute. Order Document No. 5034 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

but their favorable scores dropped and their unfavorable scores rose considerably.

Thus, within the limitations imposed by the compelling perceptual features of the Heider-Simmel movie, preinformation not only affected impressions of the figure about which it was given but also changed impressions of the other two figures. As impressions of *T*'s personality became more or less favorable, impressions of *t*, and, to a lesser extent, impressions of *c*, varied inversely. It is noteworthy that impressions of *c*'s personality changed less than did impressions of *t*. The major source of "human" interaction in the movie is seen as occurring between *T* and *t* with *c* remaining relatively passive. Since *c* interacts less with *T* than does *t* it is understandable why the different preinformation about *T* affects impressions of *c*'s personality less than it affects *t*'s.

SUMMARY

It was hypothesized that the animated geometric figures in the Heider-Simmel movie interact in such a way that the impression formed by *Ss* about the personality of one figure about which preinformation is given affects the way the other two figures are seen. This hypothesis was investigated by comparing the frequency with which favorable or unfavorable traits were ascribed to these figures under three conditions: favorable preinformation about one figure (100 *Ss*), unfavorable preinformation about this figure (100 *Ss*), and no preinformation (100 *Ss*). If it is legitimate to generalize from animated geometric figures to real people in everyday life situations, the confirmatory results obtained tend to indicate that the impressions formed of an individual may be a function of the characteristics ascribed to another person seen to interact with him.

REFERENCES

1. ASCH, S. E. Forming impressions of personality. *J. abnorm. soc. Psychol.*, 1946, **41**, 258–290.
2. BERNALES, I. The influence of family stereotypes in the judgment of personality. Unpublished master's thesis, Smith Coll., 1947.
3. GOLLIN, E. G. Forming impressions of personality, *J. Pers.*, 1954, **23**, 65–76.
4. HEIDER, F., & SIMMEL, MARIANNE An experimental study of apparent behavior. *Amer. J. Psychol.*, 1944, **57**, 243–259.
5. HERSHKOWITZ, A. A further investigation into the processes of forming impressions of personality. Unpublished master's thesis, New School of Social Research, 1954.
6. KELLEY, H. H. The warm-cold variable in first impressions of person. *J. Pers.*, 1949, **18**, 431–439.
7. LUCHINS, A. S. Forming impressions of personality: a critique. *J. abnorm. soc. Psychol.*, 1948, **43**, 318–325.
8. MENSH, I. N., & WISHNER, J. Asch on "forming impressions of personality": Further evidence. *J. Pers.*, 1947, **16**, 188–191.
9. SIGURDARDOTTIR, V. An approach to the study of stereotypes of family roles and their influence on people's judgments of personality and behavior. Unpublished master's thesis, Smith Coll., 1946.

Received October 7, 1955

THE EFFECT OF STIMULUS CLASS ON THE PROBLEM-SOLVING BEHAVIOR OF SCHIZOPHRENICS AND NORMALS

ROBERT H. DAVIS AND ROBERT W. HARRINGTON[1]

The RAND Corporation

ONE characteristic of the hospitalized schizophrenic is that generally he has been unable to achieve socially acceptable solutions to the problems of dealing with other individuals. Whether this inability is limited to certain classes of problems, i.e., those involving human stimuli, or indicates a generalized inability to solve problems, is of importance for personality theory. If for example, the inability of the schizophrenic to engage in effective problem-solving behavior is limited to problems involving human stimuli and not other classes of stimuli, it would be difficult to account for this fact with a theory of personality that stressed the importance of physiological status and failed to consider adequately the effects of interpersonal relations.

The present study was undertaken to ascertain the relative ability of schizophrenic and normal subjects (*Ss*) to engage in problem-solving behavior when the stimuli involved are of two types; human and nonhuman.

[1] The present study was conducted while the authors were at Allegheny College, Meadville, Pa., and the VA Hospital, Marion, Ind., respectively. The authors wish to express their appreciation to Mary Jane Curry and Marie C. Morane both of whom participated in the collection and organization of the data, and to the staff of the Warren State Hospital, Warren, Pa.

TABLE 1

MEAN DIFFERENCE BETWEEN NORMALS AND SCHIZOPHRENICS MATCHED ON ONE OF THE TWO CLASSES OF STIMULI AND COMPARED ON THE OTHER

Matching Variable	Comparison Variable	Matching Mean	*N* (Pairs)	Schizophrenic Mean	Normal Mean	Mean Diff.	*t*	*p* Sig. Level	Direction of Difference
Abstract	Human	3.57	45	3.25	3.78	.53	2.82	.01	N > S*
Human	Abstract	3.43	44	3.69	3.40	.29	1.40	n.s.	S > N

* N.: Normal. S: Schizophrenic

METHOD

Although not a study in information theory per se, the approach used here to the measurement of problem-solving behavior was suggested by the work of Miller (2, 3) and others (1, 4, 5). The task involved the presentation of 32 pictures from which *S* was required to choose, on the basis of information supplied by *E*, the correct one. One set of pictures was heavily weighted with human content; the other contained only pictures devoid of human content.

Subjects. Fifty patients from the Warren State Hospital, Warren, Pennsylvania, and 50 students from Meadville High School, Meadville, Pennsylvania, served as *S*s. Each patient had been diagnosed Schizophrenic Reaction and was judged by the hospital staff to be in good contact. No history of psychiatric disorder was present in the student population. Each patient was equated with a student on the basis of either IQ or grade achievement.[1]

Procedure. In both the human and nonhuman problems the pictures were selected so that when five bits of information had been given to *S*, he could locate any given picture out of the 32 arranged before him. At the outset of the study, standardized instructions explaining the test were read to *S*s, and they were encouraged to ask questions. All questions were answered fully and, if necessary, the directions were repeated. During the problem, *S*s were not told the extent to which they were correct or incorrect. Both groups were encouraged to listen closely and to do their best.

In the problem with human content, half of the pictures contained a male and half contained a female. The 16 male pictures were equally subdivided in the following manner: tie, no tie; book-in-hand, no book-in-hand; glasses, no glasses; and hat, no hat. The 16 female pictures were equally subdivided in the following manner: scarf, no scarf; book, no book; glasses, no glasses; and coat, no coat. When *S* was presented with the 32 cards containing human stimuli, he was given five bits of information, one bit at a time. Following the presentation of each bit, *S* was asked to indicate which of the 32 cards it might be. Thus, if the first bit was the statement, "It's a male," and if *S* selected a card with a male figure, he was credited with one bit. He might then be told that "the person for whom we are looking is wearing a tie." If *S* correctly indicated a male with a tie, he was credited with a second correct bit. Then he might be informed that "the person for whom we are looking is carrying a book." Proper identification of a male with a tie and a book earned an additional bit of credit. This procedure was continued through five bits, and each *S* was given a score from zero to five, depending upon the number of bits of information about the cards which he could effectively utilize. The initial bit of information was alternately male or female. All *S*s were given two trials with cards containing human content and two trials with cards containing nonhuman content.

The cards with nonhuman content were designed to parallel the human content cards and the problem was presented in the same manner. These cards, drawn with a Leroy Lettering Set, were divided into the following five bits: (*a*) printed in blue or red; (*b*) letters or nonletters; (*c*) if letters, big or little, and if nonletters, then numbers or figures; (*d*) printed on the top or bottom of the card; and (*e*) encircled or boxed in. Before presenting these cards to *S*s the difference between letters and nonletters was explained to them. Numbers and figures were pointed out and the general differences noted.

RESULTS AND DISCUSSION

For the purpose of comparing the two populations, each schizophrenic *S* was matched with a normal *S* on the basis of his performance on the nonhuman task. These *S*s were then compared by *t* tests for matched groups for performance on the problem involving human stimuli. When the groups were compared for performance on the nonhuman problem, a similar technique was used except for matching on the basis of performance on the task involving human stimuli. An inspection of Table 1 indicates that when *S*s were matched on the basis of their performance on the nonhuman problems, a significant difference ($p = .01$) was found between schizophrenic and normal *S*s in their ability to deal with the human pictures. When comparisons were drawn between the same two groups, matched for performance on human pictures, no significant difference existed between them in their ability to solve problems with nonhuman content.

It is apparent that the observed difference between normals and schizophrenics could be a function of either the stimulus class or the level of task difficulty. It is conceivable, for example, that

[1] No attempt was made to equate for age since previous exploratory studies by the authors indicated Pearsonian *r*s of very low order between performance on the task and age.

the inferior performance of schizophrenics on the human task was a function of the increased difficulty of the human task, independent of the stimulus class. There are two considerations which indicate that the crucial variable was stimulus class rather than level of problem difficulty. On the one hand, comparison of the performance of the normal *Ss* only on the two classes of stimuli indicates no significant differences with respect to mean scores of 3.57 (nonhuman) and 3.78 (human), and 3.43 (human) and 3.40 (nonhuman). Matched *t*s, computed for both sets of means, were not significant (1.08 and .15, respectively). If the two tasks differed in level of difficulty, one would expect this difference to be apparent in the performance of normal *Ss* on them. Normal *Ss* might not show such a difference, however, if both the abstract and human tasks were so easy that asymptotic scores were obtained on both. An inspection of Table 1 indicates that normal *Ss* did not obtain scores so near the upper limit (5.0) as to allow the conclusion that the tasks were exceptionally easy for either group of *Ss*.[3]

These results appear to indicate that the problem-solving ability of schizophrenics is a function of the class of stimuli with which they must deal and not necessarily a deficiency in problem-solving ability per se. Two such classes were employed in the present study. One of these classes involved human stimuli and the other nonhuman stimuli. Had the *Ss* been compared using other classes of stimuli, e.g., sexual symbols, mother figures, aggressive scenes, different racial groups, etc., the results might well have been different. In other words, a technique similar to the one used in this study might provide a method for isolating classes of stimuli which consistently and differentially disrupt or impair the problem-solving ability of psychologically distinct groups of people, e.g., normal and psychotic; prejudiced and nonprejudiced; extroverted and introverted, etc. Such information could be used to test various theories of personality and could also provide important data to the practicing clinician.

[3] Test-retest reliability coefficients were also calculated for these tasks. With respect to the human stimuli, the coefficients were .72 for normals and .66 for schizophrenics.

SUMMARY

The ability of normal and schizophrenic *Ss* to deal with information involving human and nonhuman stimuli was compared. When *Ss* were matched on the basis of ability to utilize information about nonhuman stimuli, it was found that the problem-solving behavior of schizophrenics was significantly disrupted by the presence of human stimuli in the test situation. When the same populations were matched for ability to solve problems involving human stimuli, no significant difference was found in their capacity to deal with information involving nonhuman stimuli. The importance of this unilateral disruption of problem-solving behavior for personality theory was pointed out, and it was suggested that the technique utilized in the present study could be used to explore the effect of numerous classes of stimuli on normal and schizophrenic behavior.

REFERENCES

1. MILLER, G. A., & FRICK, F. C. Statistical behavioristics and sequences of responses. *Psychol. Rev.*, 1949, **56**, 311–324.
2. MILLER, G. A. What is information measurement? *Amer. Psychologist*, 1953, **8**, 3–11.
3. MILLER, G. A. *Language and communication*. New York: McGraw-Hill, 1951.
4. ATTNEAVE, F. Some informational aspects of visual perception. *Psychol. Rev.*, 1954, **61**, 183–193.
5. SHANNON, C. E., & WEAVER, W. *The mathematical theory of communication*. Urbana: Univer. of Illinois Press, 1949.

Received February 6, 1956.

RESPONSE SET IN THE F SCALE[1]

LOREN J. CHAPMAN[2]
University of Chicago
AND DONALD T. CAMPBELL
Northwestern University

In the widely used F scale measure of authoritarian personality trends (1) all of the items are worded in a single direction, such that the "authoritarian" tends to agree with all items, the "nonauthoritarian" to disagree. Cronbach (4, 5) has marshalled considerable evidence to indicate that total scores on tests constructed in this fashion contain some systematic variance adhering not to the content of the items, but rather to a "response set" of "acquiescence," independent of specific item content. Consistent individual differences have repeatedly been found in the tendency to agree with items in general, and if not controlled in test construction, this "acquiescence" bias can generate sizable reliabilities within tests and correlations between tests that are independent of the reactions of persons to the specific content of the items. While, as Berg (2) suggests, acquiescence may in itself yield a useful personality measure, it often leads to a misinterpretation of findings when consistent individual differences are attributed to item content rather than to this more general trait. The standard precaution against such response bias in attitude research is to write items worded in both directions, so that the ideal high and low scorers must agree with half of the items and disagree with the others. The present paper reports on two efforts to so revise the F scale by reversing the wording of some F scale items.

In most attitude scales there is a specifiable social object presented, along with reactions to it, and it is a simple matter to present both favorable and unfavorable items. With the F scale the problem is much more difficult, and the present writers can make no defense of the particular reversals used. Take for example the item, "The wild sex life of the old Greeks and Romans was tame compared to some of the goings-on in this country, even in places where people might least expect it." How is this to be reversed? One could say, "The sex life that goes on today even in places where one least expects it, is nothing compared to the wild sex life of the Greeks and Romans." Thus reversed, the item would still seem to appeal to the prurient moralizing that the initial item sought to tap. Better perhaps to state it, "Neither was the sex life of the Greeks and Romans, nor is the sex life of today as wild and immoral as is commonly thought." In the end we decided to leave this item alone.

In the first study we employed ten items in their original form[3] and ten reversed items. Because these reversals are an essential part of the evidence, they are presented in full as the first 10 items in Table 1, which also indicates the test form and item number of the original item in *The Authoritarian Personality* (1) from which each was modified. In some analyses these two sets of items are employed as one 20-item test, but more frequently they are treated as separate 10-item tests, with the original wording being designated as F+, the reversals as F−. All items have been scored in the authoritarian direction.

In the first study, these 20 items were included in a longer questionnaire and administered to 134 students (69 male, 65 female) from the introductory course in psychology at Northwestern University, fall, 1953. Five levels of response were offered: strongly agree, agree, undecided, disagree, strongly disagree, but for scoring purposes the responses were dichotomized as agree and disagree, the undecided responses on each item being included with the less frequently chosen of the two. The F+ scale had a reliability[4] of .69, the F− scale a reliability of .41. The correlation between them was positive and in the proper direction, but was only .17, significant at but the .05 level. To explore this failure in item reversal at the item-by-item level, tetrachoric correlations were computed between the F− items and the F+ total score. These values are also shown in Table 1. Note that the highest value, −.30, is in the wrong direction, indicating that the high-F+ person tends to agree with the item. As checked by χ^2 this relationship is significant at the .03 level. Whereas in the

[1] This study was supported in part by the United States Air Force under contract No. AF 18(600)-170 monitored by the Crew Research Laboratory, Air Force Personnel and Training Research Center, Randolph Air Force Base, Randolph Field, Texas. Permission is granted for reproduction, translation, publication, use and disposal in whole and in part by or for the United State Government.

[2] This manuscript was written while the senior author was a post-doctoral research fellow of the National Institute of Mental Health.

[3] These were items 1, 13, 19, 26, 31, 37, 38, 42 of Form 40, and 55 and 67 of Form 78.

[4] All reliabilities were computed by an internal consistency method of Gulliksen (6).

TABLE 1

THE F− ITEMS AND THEIR CORRELATIONS WITH SCORE ON THE F+ SCALE

F+ Version*		F− Version	First Testing	Second Testing
Form	Item			
40	9	When a person has a problem or worry, he should try to think it through, rather than try to forget by keeping busy with cheerful things.	−.30	
40	2	It takes more than will power to conquer most obstacles.	−.11	
60	44	It interferes with learning and effective work when our teachers or bosses outline in detail what is to be done and exactly how to go about it.	−.10	
40	27	It would be better if people admitted to themselves that they do not always feel a great love, gratitude, and respect for their parents.	.12	.07
40	6	In spite of what some people say, human nature does not make war and conflict inevitable.	.18	−.06
40	21	It would be a good thing if more young people kept their youthful, rebellious ideas as they grew older.	.18	−.12
40	23	Strong leaders in whom people put complete faith are not what our country needs today.	.09	.12
40	34	Getting rid of all of the immoral, feebleminded and crooked people would not actually solve many of the country's problems.	.28	−.06
40	41	The artist, writer and philosopher are probably more important to society than the business man and manufacturer.	.28	.10
40	43	Familiarity breeds respect more often than it breeds contempt.	.15	−.04
40	2	With most men it is luck and help from others rather than will power and drive that determines what eminence they attain in life.		−.16
40	8	It is unnecessary and undesirable for a person to have complete faith in some supernatural power whose decisions he obeys without question.		.06
40	12	Even though a person has bad manners, habits, and breeding, he should still be able to make himself acceptable to other people.		.02
40	16	It is completely absurd to believe that some people may be born with an urge to jump from high places.		−.03
40	18	Even though nowadays many different kinds of people move around and mix together a great deal, it is foolish to worry much about catching an infection or disease from them.		−.16
60	44	It interferes with learning and effective work when our teachers and bosses assume too active a leadership and tell us exactly what to do.		−.20
78	24	There have been other ages with more instability, conflict, and imminent danger than ours.		.07
78	38	It is a good thing to maintain the present emphasis in colleges on intellectual and theoretical topics, rather than change the emphasis to practical matters.		−.03

* For original items, see (1).

original wording the high-F person was anti-introceptive (When a person has a problem or worry, it is best for him not to think about it, but to keep busy with more cheerful things), by agreeing with the reversed wording (first item, Table 1) he appears introceptive. What is in common between the two wordings is that both are plausible declarative statements. Items 34 and 41 dealing with getting rid of the immoral and feebleminded, and the value of artists vs. businessmen, seem to have been more successfully reversed, and are now in a form in which the high-F person tends to disagree with them.

In the second effort at reversing F scale items, 16 F+[5] and 15 F− items were employed. The F− items included the 7 of the first set which showed positive correlations, plus 8 new reversal attempts. These items are also shown in Table 1. These 31 items along with others were included in a questionnaire administered to 144 male students from introductory psychology courses at Northwestern University in the fall of 1954. The reliability of the F+ items was .53. The reliability of the F− items was .41. The correlation between the two scales was −.01. Again, tetrachoric correlations between

[5] These were the 10 items used in the first study, plus Nos. 4, 9, 25, 33, 35, and 39 of Form 40.

the F+ total and the F − items were computed. As can be seen from Table 1, the values are low and predominately negative.[6] Although we have explored but a small sample of the ways in which these and other F scale items could be reversed, it seems fair to conclude that the achievement of a successful reversal would be a very difficult task.

It was possible in the second study to obtain a rather pure measure of the acquiescence response set through a special scoring of other questionnaire items administered at the same time. These items were intended to measure (*a*) superior orientation or a tendency to identify with one's superiors rather than with one's subordinates in situations involving cross pressures from these echelons; (*b*) alienation, or a tendency toward cynicism about the sincerity of other persons and conventional values; (*c*) identification with disciplining authority; and (*d*) preference for cooperation and doing things in groups. For each of these topics, an equal number of positive and negative items were constructed, to balance out the acquiescence response set. Rescoring the items for a general tendency to agree should thus produce a measure of the response set in which content factors were balanced out. Based on 74 items, this measure showed a reliability of .63. It correlated .43 with F+, −.06 with F−, and .32 with total F.[7]

DISCUSSION

If the F scale were free of response-set variance, one would expect F+ and F− to correlate positively, to a degree approaching their reliabilities. On the other hand, if the systematic variance picked up by the F scale scores were only acquiescence response set, then F+ and F− should show a high negative correlation. The zero or low positive correlation found indicates that neither extreme is true, and that the typical F scale score probably contains both response-set variance and variance adhering to the content of the items.

But without rejecting this conclusion we can note that it does not suffice for the data presented. On this basis taken simply, F− should correlate as highly (though negatively) with the pure response-set measure as does F+. It has about the same number of items, and all items are worded in one direction. But this outcome does not appear. It seems necessary to consider that the acquiescence response-set variance is correlated positively with the item-content variance of the F scale. The trait of agreeing with items seems to be symptomatic of a more general personality syndrome also symptomized by belief in the F scale content measured independently of the response set. Our best evidence of this is the highly significant correlation of .32 between the pure response-set measure and Total F in which response-set variance is presumably canceled out by the combination of items worded in both directions. On this basis, we would interpret the correlation of .43 between the response-set measure and F+ as due to the summation of the effects of response set in the F+ items plus content in the F+ items. The F− items are so worded that these two sources of variance each potentially correlated with the pure set measure, subtract from each other and cancel each other out, resulting in the −.06 value. This interpretation provides in addition a rationale for the lower reliabilities of the F− scales.

This interpretation by no means implies that the F scale will correlate only with test scores containing the response set. Ample evidence to the contrary is available, in which the F scale correlates significantly, if not highly, with measures of quite different construction (3). However, it does imply that the correlations with tests similarly constructed will be higher than they would be were the response-set bias controlled as by having items worded in both directions, by multiple choice format, or otherwise. This observation has important implications for some of the absolute levels of correlation reported in the initial study (1). Like the F scale, all of the Ethnocentrism items were worded in the same direction; thus the acquiescence response set must raise the correlation above what it would otherwise be. For the Political and Economic Conservatism scale, on the other hand, items worded in both the conservative and the liberal directions were employed. In this fashion, as well as in terms of the smaller number of items used, the dice were inadvertently loaded against finding political-economic attitude as important a correlate of ethnocentrism as the F scale measure of personality.[8]

[6] The difference between the two studies in levels of correlation found may be due to the fact that the second study was based on males alone. In general, in our experience in this area, correlations and reliabilities run higher for women than for men.

[7] In a further exploration, these items were split into 40 worded in the first person ("I feel," etc.) and 34 worded as declarative statements, as in the F scale. The acquiescence score for these had reliabilities of .46 and .47 respectively, correlated with each other .45, and correlated with F+ .41 and .33. The difference in F+ correlations was not significant by the Pearson and Filon test (7), and was not in the expected direction.

[8] Since this paper was prepared, two other F scale item-reversal studies have appeared: Leavitt, H. J., Hax, H., and Roche, J. H. "Authoritarianism" and agreement with things authoritative. *J. Psychol.*, 1955, **40**, 215–221, and Bass, B. M. Authoritarianism or acquiescence, this Journal, 1955, **51**, 616–623. Both

SUMMARY

An attempt was made to evaluate acquiescence response-set variance in the F scale measure of authoritarian personality trends by employing reversed wordings of items. Response set is found to be an important factor in the F scale which cannot easily be corrected by the practice usual in attitude-test construction of employing both positively and negatively worded items. Evidence seems to indicate in addition that the content component of the F scale is correlated with the acquiescence response set.

concur in finding evidence of response set variance in the ordinary form of the F scale. Bass confirms the lower F− reliability found here. Leavitt does not report reliabilities. Neither study provides the correlations of both F+ and F− with an external measure of response set which provides the major interpretive argument of the present paper.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J. & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. BERG, I. A. Response bias and personality: the deviation hypothesis. *J. Psychol.*, 1955, **40**, 61–72.
3. CHRISTIE, R. Authoritarianism re-examined. In R. R. Christie & Marie Jahoda (Eds.), Studies in the scope and method of "The Authoritarian Personality." Glencoe, Ill.: Free Press, 1954. Pp. 123–196.
4. CRONBACH, L. J. Response sets and test validity. *Educ. psychol. Measmt*, 1946, **6**, 475–494.
5. CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt*, 1950, **10**, 3–31.
6. GULLIKSEN, H. *Theory of mental tests*. New York: Wiley, 1950.
7. PETERS, C. C., & VAN VOORHIS, W. R. *Statistical procedures and their mathematical bases*. New York: McGraw-Hill, 1940.

Received February 15, 1956.

A NOTE ON "ETHNOCENTRISM" AND ACQUIESCENT RESPONSE SETS

DOUGLAS N. JACKSON AND SAMUEL J. MESSICK[1]

The Menninger Foundation

BOTH the California E scale and F scale (1) are composed entirely of positively stated items. The items are so worded that agreement with them indicates "ethnocentric" or "authoritarian" attitudes while disagreement presumably indicates the "equalitarian" end of assumed unidimensional scales. Recent investigations (2, 7, 8) have emphasized the importance of such a positive item form in using the F scale as a measure of attitude. These studies assessed the relationships between F-scale items in their original positively-stated form and F-scale items rewritten in "reversed" form so that disagreement rather than agreement could be considered an indication of authoritarian attitudes. The correlations between agreement to reversed and original items were generally found to be *positive*, i.e., people who agreed with positively-stated F-scale items also tended to agree with the reversed items even though the item content was opposite in direction. Thus, these studies indicate that when the accepted test practice of counterbalancing the direction of item content (9) is used in specifying an authoritarian scale, a paradoxical property of *negative reliability* is obtained.

This evidence obtained with the F scale may also be interpreted in terms of what Cronbach (4) has called an acquiescent response set, which has implications for interpreting the reported high correlations between the F scale and the E scale. There is a possibility that the observed covariation of responses to the two scales may be in part attributable to a consistent generalized tendency to agree or acquiesce to items in print, rather than being solely a function of consistent ideological beliefs.

In the present study, 93 junior and senior students[2] in high school, 40 males and 53 females ranging in age from 16 to 18, were given the 9-item E scale, Forms 40 and 45 (1), the reversed F scale,[3] and the Gough intolerance scale (6). The groups were selected from social studies classes and contained only white, Christian students.

The reversed F scale was constructed by rewriting each item so that the content would appear to reflect a viewpoint opposed to the original, while retaining a similar style of expression. For

[1] This study was completed while the authors were USPHS Postdoctoral Clinical Research Fellows of the National Institute of Mental Health. Dr. Jackson is now at the Pennsylvania State University and Dr. Messick is at the Educational Testing Service and Princeton University.

[2] The writers are indebted to Mr. E. B. Weaver, Principal, Topeka High School, for kindly consenting to cooperate in this project.

[3] The authors wish to thank Dr. Irwin Miller and James E. Simpson for help in reversing the items.

TABLE 1
THE REVERSED F SCALE

1. A love of freedom and complete independence are the most important virtues children should learn.
2. All the will power in the world will not help us when weaknesses and difficulties stand in our way.
3. The success of modern science proves that every important thing can ultimately be understood by the human mind.
4. Because human nature is improving, war and conflict will eventually be eliminated.
5. Every person should have complete faith in his own independent judgment, not in some supernatural power whose decisions he obeys without question.
6. When a person has a problem or worry, he should drop everything and concentrate upon it until the solution appears.
7. The deeper and more enduring qualities in getting along well with people are far more important than external things like manners, habits, and breeding.
8. What the youth needs most is complete freedom from discipline in order to flow naturally into the onward stream of cooperative progress for the benefit of all mankind.
9. It is known with complete certainty that the urge to jump from high places is learned, not inborn.
10. Nowadays, since democracy demands that people of widely different background and station mix together, a person should not be finicky about catching a disease from any of them.
11. An insult to our honor should always be overlooked, for "whosoever shall smite thee on thy right cheek, turn to him the other also."
12. The rebellious ideas that young people sometimes get must be encouraged and developed at all costs to guarantee mature citizenship in adulthood.
13. Only if prewar authorities are kept out of the German government, will true democracy be achieved in that country.
14. What this country needs most, more than willful leaders, are laws and political programs requiring all citizens to actively share responsibility.
15. Sex offenses, such as rape and attacks on children, never merit punishment; such offenders should always be treated with kindness and sympathy by qualified psychiatrists.
16. All attempts to divide people into the two distinct classes of the weak and the strong are doomed to failure.
17. Every truly mature person outgrows childish feelings of submissive respect and of excessive love and gratitude for his parents.
18. Such pseudosciences as astrology have never explained anything and never will.
19. The surest way to insure the disappearance of the true American way of life is to resort to the use of force.
20. In this scientific age there can be no justification for denying investigators the right to study so-called personal and private matters.
21. The complete end to wars and social problems will soon be realized.
22. There would be no immoral, crooked, or feebleminded people, if we could get down to brass tacks and somehow get rid of our social problems.
23. The wild sex life of the old Greeks and Romans makes sexual goings-on in this country seem tame.
24. If people talked things over and didn't work so much, everybody would be better off.
25. It is foolish and ridiculous to have ideas that our lives could possibly be controlled by plots hatched in secret places.
26. Homosexuals are never criminals and must not be punished as such.
27. The business man and the manufacturer are not any more important to society than the artist and the professor.
28. No sane, normal, decent person can avoid having thoughts of hurting close friends or relatives.
29. We are bound to admire and respect a person if we get to know him well.
30. Nobody ever learned anything really important through suffering.

example, "Obedience and respect for authority are the most important virtues children should learn," was modified to read "A love of freedom and complete independence are the most important virtues children should learn," and "No weakness or difficulty can hold us back if we have enough will power," was changed to "All the will power in the world will not help us when weaknesses and difficulties stand in our way," etc. (see Table 1). There are probably alternative items that would better represent the "reverse" of the original statements. However, a defense of the adequacy of the reversals used in the present study would seem largely irrelevant if a certain typology like "authoritarianism" is to be defined in terms of belief in a particular ideological content and presumably disagreement with its reverse.

The Gough intolerance scale is a group of items empirically derived from the MMPI using the differentiation of extreme groups on the Levinson-Sanford Anti-Semitism scale as a criterion. The intolerance scale has been shown to correlate significantly with the California F scale, .46 and .43 (6). Incidentally, it should be noted with respect to response sets that 29 of the 32 items on Gough's scale are scored "true" to be in the "intolerant" direction. Gough (6) also reports a high negative correlation (−.73) between the intolerance scale and the MMPI *K* scale, a measure of test-taking attitude that contains 29 of 30 items scored false.

Cohn (3) has found a significant correlation between the F scale and a tendency to answer "true" to a diverse group of MMPI items and has

suggested that the correlation was due to a generalized tendency to agree with statements. Since the preponderance of items scored "true" on the intolerance scale suggests that it too may indicate some generalized tendency to acquiesce, we might expect such a response tendency to be reflected in a positive correlation between agreements to intolerance-scale and reversed F-scale items, even though there is a positive correlation between the intolerance scale and the original F scale. Of course, a positive correlation would be expected between the E scale and Gough's intolerance scale, because of their common empirical ancestry.

The results indicate that even though the Gough intolerance scale correlated positively with the E scale ($r = .24$, $p < .02$) in our sample, it also correlated *positively* with agreement to reversed *F*-scale items ($r = .23$). Thus, subjects who agreed with the intolerance-scale items tended to agree with reversed F-scale items. However, the fact that the intolerance scale appears to reflect an acquiescent response set may even contribute to its construct validity (5), but extreme caution should be exercised in defining the characteristics of respondents on the basis of item content. Also, the correlation between the E scale and the reversed F scale was .03, rather than being significantly negative as would be expected from item content. These results highlight the necessity of controlling item form in studying certain attitude contents, such as "ethnocentrism" or "authoritarianism." Also, previously reported relationships between beliefs and personality may take a different form when more adequate attitude measures are used.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. BASS, B. M. Authoritarianism or acquiescence? *J. abnorm. soc. Psychol.*, 1955, **51**, 616–623.
3. COHN, T. S. The relation of the F scale to a response set to answer positively. *Amer. Psychologist*, 1953, **8**, 335.
4. CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt*, 1950, **10**, 3–31.
5. CRONBACH, L. J., & MEEHL, P. E. Construct validity in psychological testing. *Psychol. Bull.*, 1955, **52**, 281–302.
6. GOUGH, H. G. Studies of social intolerance: II. A personality scale for anti-Semitism; III. Relationship of the *Pr* scale to other variables. *J. soc. Psychol.*, 1951, **33**, 247–262.
7. JACKSON, D. N. Stability in resistance to field forces. Unpublished doctor's dissertation, Purdue Univer., 1955.
8. LEAVITT, H. J., HAX, H., & ROCHE, J. H. "Authoritarianism" and agreement with things authoritative. *J. Psychol.*, 1955, **40**, 215–221.
9. LOEVINGER, JANE. Some principles of personality measurement. *Educ. psychol. Measmt*, 1955, **15**, 3–17.

Received February 17, 1956.

A NOTE ON "DISCRIMINATION REACTION TIME AS A FUNCTION OF ANXIETY AND INTELLIGENCE"

JAMES T. FREEMAN AND HOWARD MAHER

Iowa State College

GRICE (3) has recently reported a functional relationship between intelligence and discrimination reaction time in accounting for results otherwise attributable to the operation of anxiety as a drive variable. It appears that the rationale for such an interpretation rests, to a great extent, upon his assertion that "... the discrimination reaction time test is known to be correlated with intellectual factors ..." (4). This statement does not appear to be completely in accord with the facts as reported in Grice's reference (4) and elsewhere (1).

On examination of Melton's data (4, pp. 1027–1030), one finds that discrimination reaction time (DRT) is only moderately related to an "intellectual" factor. Further, beyond the first trial in DRT performance, intellectual components virtually disappear. On the other hand, such matters as perceptual speed and simple reaction time assume somewhat greater importance as performance progresses (1). Consequently any test involving perceptual factors may be expected to predict a perceptual criterion (DRT performance) reasonably well. Grice reported that the Clerical Aptitude Test, which he used as a "general intellectual measure," involves "perceptual factors" among others. Guilford (2) indicates that such "perceptual factors" correlate about as well as any others in predicting DRT performance.

In short, it would seem that Grice essentially demonstrated that a psychometric device with at least some weighting of perceptual factors does

reasonably well in predicting a type of performance that is largely perceptual in character! One may therefore question whether a functional relationship between *intelligence* and DRT has been reliably indicated and whether such a relationship is substantively likely. Finally, one may question the use of the Clerical Aptitude Test as the basis for equating groups for intelligence. Were not Air Force data on tests more universally recognized as general intellectual measures available?

It is believed that some of the points raised in the present note should have additional significance in view of the increasing use now being made of psychometric measures in the prediction and interpretation of experimental samples of behavior. There is implied in such an approach a joint responsibility among several research areas.

REFERENCES

1. FLEISHMAN, E. A., & HEMPEL, W. E., JR. The relation between abilities and improvement with practice in a visual discrimination reaction task. *J. exp. Psychol.*, 1955, **49**, 301–312.
2. GUILFORD, J. P., & LACEY, J. I. (Eds.) Printed classification tests. Washington: U. S. Government Printing Office, 1947. (*AAF Aviat. Psychol. Program Res. Rep.* No. 5.)
3. GRICE, G. R. Discrimination reaction time as a function of anxiety and intelligence. *J. abnorm. soc. Psychol.*, 1955, **50**, 71–74.
4. MELTON, A. W. (Ed.) Apparatus tests. Washington: U. S. Government Printing Office, 1947. (*AAF Aviat. Psychol. Program Res. Rep.* No. 4.)

Received February 27, 1956.

SOCIAL STATUS AND ABSTRACT BEHAVIOR

DONALD C. FINDLAY[1] AND CARSON McGUIRE

Human Research Unit No. 1
Fort Knox, Kentucky

University of Texas

IN recent years "intelligence tests" have been sharply criticized for an alleged cultural bias. It has been argued that this bias handicaps the bright child who has grown up in a lower socioeconomic environment. The present-day intelligence test, according to the argument, places a premium on knowledge readily accessible to the child in a middle-class home but seldom accessible to the child in a lower-class home.

Thus, when a lower-class child and a middle-class child are found to have similar IQs, it is possible that the lower-class child is actually "genetically" brighter, and that his performance was handicapped by the cultural bias of the test (3, p. 4). Presumably, if the two children were tested with intellectual problems involving materials with which they were equally familiar (2, p. 66), the lower-class child would probably make a higher score.

The present study is an attempt to evaluate this assumption by comparing children of similar IQ but of dissimilar socioeconomic background on a series of sorting problems requiring abstract behavior. The problems appeared to involve a knowledge of only four relatively simple block characteristics: height, width, shape, and color.

The hypothesis tested was: If lower-status children and middle-status children of similar IQ are given block-sorting problems involving concepts of equal familiarity to the two groups, the performance of the lower status children will be significantly superior to that of the middle-status children.

[1] This study was adapted from a doctoral dissertation submitted to The University of Texas by the senior author.

METHOD

Subjects. Ninety-six subjects, selected by age-grade, index of status characteristics (ISC), and intelligence quotient, were used. Thirty-two *S*s with IQs ranging from 90 to 110 were selected nonsystematically from each of three grades (the third, sixth, and ninth) in the public school system of a central Texas community. Half of the *S*s at each grade level were selected from family backgrounds which rated as lower status (51 or less) by ISC standards. The other half of the *S*s at each grade level were selected from middle-status (higher than 51) family backgrounds.

Age-grade refers to children of the same age at a given school-grade level. The index of status characteristics is the social class index developed by Warner and his associates (5). A modified form of the ISC based on occupation, dwelling area, and source of income was used (4). IQ was determined by means of the California Test of Mental Maturity. Third-grade *S*s were tested on the Primary Short-Form, sixth-grade *S*s on the Elementary Short-Form, and ninth-grade *S*s on the Intermediate Short-Form. Differences between mean IQs of the two status groups at each age-grade were not significant (t's of .11, .18, .13).

Materials. Four block-sorting problems similar to those of Ach (1) were used. In each problem, the *S*'s task was to sort 20 blocks into four five-block groups.

The four problems were derived from a matrix of

blocks representing combinations of four characteristics: height, breadth, color, and shape. There were two values of height (short and tall) and breadth (wide and narrow), and three values of color (blue, red, yellow) and shape (round, square, hexagonal). Blocks for a given problem were selected from the matrix so that only one sorting solution (basis of classification) was possible with the available values of the four characteristics.

The four sorting solutions, the four bases of classification, were always combinations of two characteristics (such as shape and breadth), each of which was present in two values (such as round and square) in the particular problem. In order to sort correctly, the *S* had to look at the 20 blocks and "abstract" the two characteristics which would permit a fourfold (2 × 2) classification. For example, in the problem based on shape and breadth, the correct solution consisted of sorting the blocks as follows: five round-wide blocks, five round-narrow blocks, five square-wide blocks, and five square-narrow blocks.

The other three problems (in addition to the shape-breadth [SB] problem) had solutions based on shape-height [SH], color-breadth [CB], and color-height [CH], respectively.

Procedure. The experiment was conducted in two phases: (*a*) by administering three block problems to an *S*; and (*b*) by checking the *S*'s knowledge of the simple concepts on which those problems were based.

In the phase *a*, the problems were presented in four orders: CB-CH-SB, CH-SH-CB, SB-CB-SH, SH-SB-CH. In each of the three age-grades, each of the four orders of presentation was given to four *S*s from each of the two socioeconomic levels.

Each problem was presented by pointing out the various characteristics and values of the blocks to the *S*, and then instructing him to sort out four "kinds" of blocks from the 20 before him. He was told that if his sorting were correct there would be an equal number of blocks of each kind. The instructions were phrased as simply as possible and in a manner intended to establish a "game" type of atmosphere. Performance was scored by using a system of "aids" which controlled the difficulty level of the problem.

An aid consisted of showing the *S* two blocks which illustrated the solution of the problem. The first aid presented the problem. The *S* had to note the characteristics in which the two blocks were similar, and then sort the rest of the blocks into the four categories on that basis. If the problem was not solved after a specified time, a second aid representing a second sorting category of the problem was used. The procedure was repeated until all blocks had been used as aids and the problem was automatically solved. An *S*'s score was the number of blocks which he needed as aids. Thus, the larger the *S*'s score, the poorer his performance.

Phase *b*, in which knowledge of the concepts used in the block-sorting problems was checked, took place immediately after the *S*'s response to the third problem. The procedure was to spread the blocks before him, hold up a block, and ask him to pick out all blocks which were the same as this block in the two characteristics on which one of his problems was based; i.e., "Give me all the blocks that are like this one in shape and height." The *S* was then given credit for knowledge of the concept if he picked out all the correct blocks. The same procedure was followed in checking the concepts used in the other two problems.

TABLE 1

SUMMARIZED ANALYSIS OF VARIANCE OF PERFORMANCE ON BLOCK-SORTING PROBLEMS

Sources of Variance	*df*	Mean Square	*F**	*p*
Order of presentation	3	1.66	2.72	—
Age-grade	2	5.26	8.62	<.001
Social status	1	5.21	8.54	.005
OP × AG	6	.96	1.57	—
OP × SS	3	.54	.88	—
AG × SS	2	.28	.46	—
AG × SS × OP	6	.46	.75	—
Within	72	.61		

* Within-group variance used as error term.

RESULTS AND CONCLUSION

In order to determine the significance of the performance differences on Phase *a*, the block-sorting phase, between middle and lower status *S*s, analysis-of-variance procedures were used. Comparisons were made by age-grades III, VI, and IX, and by the four orders of problem presentation (Table 1). The mean difference between middle-status *S*s (12.4) and lower-status *S*s (15.2) was significant at the .005 level. The decrements in mean score from 17.1 for age-grade III, 13.1 for age-grade VI and 11.2 for age-grade IX were found to be significant beyond the .001 level. Differences between orders of presentation were not significant, nor were any of the interactions.

In Phase *b* of the experiment the chi-square test was used to compare differences between low- and middle-status *S*s' grasp of the simple concepts used in the sorting problems. None of the differences was significant. Frequencies of responses by social status (age-grade combined) for each concept yielded the following chi squares with 1 *df*: SB, .02; SH, .07; CB, 1.17; CH, .01. (Significance at the .05 level requires a value of 3.8.)

The results of the Phase *b* confirm the assumption that the lower- and middle-status *S*s had about the same grasp of the concepts used in the block-sorting problems. Nevertheless, the hypothesis that lower-status children would excel middle-status children on the sorting problems was not substantiated. Instead, it was found that the middle-status children were significantly superior.

DISCUSSION

This study does not support the argument that intelligence-type testing with "equally familiar" materials will expose the middle-class bias of standard intelligence tests. The two status groups here were equated in IQ as well as in the experi-

mental testing materials, yet middle-status *Ss* were clearly superior to lower-status *Ss*. Generalization from these findings is limited by the method of sampling (selection) used in obtaining *Ss*, as well as by the fact that only one intelligence test was represented.

The study does seem, incidentally, to support the belief that the ability to engage in abstract behavior increases with age and education through childhood and early adolescence. It is possible, however, that this finding resulted from successively more abstract content in the primary, elementary, and intermediate forms of the intelligence test used at the third, sixth, and ninth age-grades respectively to equate the two status groups.

Summary

The contention that "intelligence" tests are biased against lower-class children was investigated by testing lower- and middle-status children of similar IQs with block-sorting (abstract behavior) problems. These problems involved only four relatively simple concepts found to be equally familiar to the two groups. It was hypothesized that lower-status subjects are actually brighter than their IQs indicate, and that their performance on the block-sorting problems would be superior to that of middle-status *Ss* of similar IQ. The hypothesis was not confirmed; middle-status *Ss* were significantly superior to lower-status *Ss*.

References

1. Ach, N. *Uber die Beguffsbildung.* Bamberg: Buchner, 1921.
2. Davis, A. *Social-class influences upon learning.* Cambridge: Harvard University Press, 1950.
3. Eells, K., *et al. Intelligence and cultural differences.* Chicago: Univer. of Chicago Press, 1951.
4. McGuire, C., & White, G. *The measurement of social status,* Research paper in human development no. 3 (revised). Austin: The Univer. of Texas, May, 1952.
5. Warner, W. L., *et al. Social class in America, a manual of procedure for the measurement of social status.* Chicago: Science Research Associates, Inc., 1949.

Received March 5, 1956.

HOW "RIGID" IS THE "AUTHORITARIAN"?[1]

DOUGLAS N. JACKSON, SAMUEL J. MESSICK, AND CHARLES M. SOLLEY[2]

The Menninger Foundation

Ever since Rokeach (19) reported a relationship between "ethnocentrism" and "rigidity," a number of attempts have been made to demonstrate connections between ideological beliefs and personality variables. However, such theoretical relationships between ideology and personality are contingent upon the adequacy of the operational measures involved. The present authors question the validity of some of these measures, such as the F scale, for which the underlying dimension is defined in only one direction, i.e., scales for which agreements to positively-stated items designate one end of a continuum and disagreements with the same items presumably designate the other. The present research is an attempt to appraise the effect of such a scale form on previously observed relationships involving "authoritarianism" and "rigidity."

As a parallel to Rokeach's study, Brown (3) substituted the California F scale for the E scale and found it necessary to introduce achievement-oriented conditions before it was possible to obtain a significant relationship between Einstellung rigidity and authoritarianism. French (8), however, did not obtain these results on a military population. The instability of this relationship has led to further investigation of the properties of the measures.

Both the validity and the generality of the Einstellung test as a measure of rigidity have been questioned (8, 9, 15, 18). Levitt and Zelen (15) suggest that the consistent use of set solutions in Einstellung problems may reflect conformity and overgeneralization. Pitcher and Stacey (18) concluded that submission may be the higher order construct under which rigidity can be subsumed. It may be possible to go one step further and relate the tendency toward conformity and overgeneralization elicited by the Einstellung situation to tendencies toward conformity revealed by agreement with certain types of positively-stated items, independent of content. Such a generalized tendency to acquiesce is a source of variation to which

[1] Portions of this paper were read at the Southwestern Psychological Association meetings in Dallas, 1956.

[2] This research was completed while the authors were USPHS Postdoctoral Clinical Research Fellows of the National Institute of Mental Health. Dr. Jackson is now at the Pennsylvania State University and Dr. Messick is presently at the Educational Testing Service, Princeton. The authors wish to thank Dr. Milton Rokeach for some thoughtful comments concerning the manuscript.

Cronbach (6, 7) has repeatedly called attention under the label of "response set."

Cohn (4), noting that all the items in the California scales are unidirectional, found a significant correlation between the F scale and a tendency to answer "true" to a diverse group of MMPI items. Jackson and Miller,[3] Jackson (12), Bass (2), and Leavitt, Hax, and Roche (14), independently modified the direction of F-scale items so that disagreement rather than agreement would reflect the presumed authoritarian ideology. These studies yielded generally *positive* correlations between agreements with the original and reversed items. Such correlations do not indicate much about authoritarian ideology, but they do tell us something about the items themselves and suggest the use of a concept like negative reliability.

If this generalized tendency to agree with items independent of their content were adequately controlled, many of the relationships previously obtained with the F scale might be different (13). The present study is an attempt to interpret the reported correlation between the Einstellung test and the F scale as an indication that both measures are reflecting tendencies toward acquiescence, conformity, and overgeneralization, rather than as an indication that "authoritarians" should be labeled "rigid." Reversed F-scale items are also expected to reflect tendencies toward conformity and overgeneralization, such that the reversed F scale should correlate positively with the Einstellung test.

METHOD

Subjects

The *Ss* were 77 students from introductory psychology and history classes in a midwestern municipal university.[4] Their median age was 20, their mean age was 23.5 and the *SD* 7.6 years.

Materials

The instruments used were (*a*) the Einstellung arithmetic problems, (*b*) California F scale, and (*c*) a specially constructed reversed F scale.

The Einstellung arithmetic problems. These problems were identical with those used by Brown (3). They consisted of two examples and seven items, each of which involved withdrawing certain amounts of water from a large jar in order to obtain a desired quantity. The second example and the two "set" problems could be solved by the formula B − A − 2C, where B is the largest jar and C the smallest jar. Following these were five problems which could be solved either by the longer formula above or by the shorter formula A − C. Brown noted that this procedure elicited "long" solutions more frequently than the one used by Rokeach (19).

The original and reversed F scales. In addition to the 30 item California F scale, Forms 40 and 45 (1), *Ss* were administered a reversed F scale,[5] which was constructed by reversing the direction of the item content while retaining as much of the original style as possible. The reversed F-scale items are reported by Jackson and Messick (13).

Procedure

Two weeks prior to the administration of the Einstellung test and the reversed F scale, approximately half the *Ss* were given the original F scale. About two weeks after taking the Einstellung test and the reversed F scale, the original F scale was administered to most of the remaining half.

Since Brown observed that an ego-involving atmosphere was essential for obtaining correlations between "authoritarianism" and "rigidity," a modification of McClelland's procedure (16) was used to create achievement conditions. The experimenter, dressed in a dark suit, was introduced to the class as a psychologist who was going to administer some very important tests. Instructions were then read which suggested that the tests were measures of intelligence and leadership potential and that the *Ss* were being compared with students from a nearby state university. They were instructed to write their names on the test papers and then to proceed with the Einstellung test. After the forms were completed, they were given some information about the true nature of the test in order to relieve any anxiety that may have arisen as a result of the achievement-oriented conditions.

RESULTS

Of the 77 *Ss* to whom the Einstellung test was administered, ten made one or more errors in arithmetic and were therefore eliminated from the sample. Of the 67 *Ss* remaining, 13 who failed to use either the set or the critical solution consistently were also eliminated. This left 32 *Ss* who always used the long solution and 22 who always used the short solution.

Two single classification analyses of variance were computed between these two Einstellung groups, one comparing them with respect to original F-scale scores and the other with respect to reversed F-scale scores. Since directional hypotheses were made, one-tail tests were used. Table 1 indicates that both the F scale and the reversed F scale differentiate the Einstellung groups at the .05 point. Thus, *Ss* who agree with F-scale items tend to use the set solution on the Einstellung test. However, agreement with reversed F-scale items for the same *Ss* also tends to covary with the use

[3] Unpublished study, Purdue University, 1953. Discussed in (13).

[4] The authors are indebted to Dr. Bernard Spilka for making the F-scale data available and to Drs. Spilka and Homer Rutherford for providing class time.

[5] We wish to thank Dr. Irwin Miller and James E. Simpson for help in constructing this scale. For the 42 *Ss* taking the original F scale, mean item score was 3.10; *SD*, .67. The 54 *Ss* taking the reversed scale received a mean item score of 4.05; *SD*, .46.

TABLE 1
Differences Between Einstellung Groups with Respect to Original and Reversed F-Scale Scores

F-Scale Scores					Reversed F-Scale Scores				
Source	*df*	Mean square	*F*	*p*	Source	*df*	Mean square	*F*	*p*
"Long" vs. "Short" Solvers	1	1227	3.11	<.05	"Long" vs. "Short" Solvers	1	539	2.84	<.05
Within	40	394			Within	52	190		

of the long Einstellung solution. Thus, by changing the direction of the item content, one may obtain a "rigid" performance from people giving either "authoritarian" or "nonauthoritarian" responses.

The correlation between agreement with original and reversed F-scale items was +.35, rather than being significantly negative as consistent responses to content would require. The corrected split-half reliability of the reversed F scale was .77. The counterbalanced order of presentation of the original and reversed F scales contributed no significant effects. Also, the Einstellung results for both the F scale and reversed F scale were in the predicted direction for each of the two orders of presentation.

Discussion

The results emphasize the importance of controlling for response sets in constructing attitude scales. It seems reasonable to assume that agreement with a statement can be interpreted as reflecting belief in its content only if disagreement with the opposite content can also be obtained. The tendency to agree with an item, *any* item, especially when written in a cliché-like fashion may reflect important personality characteristics. These in turn may manifest themselves by increasing the correlational validity of the scale but not necessarily the validity as suggested by the content of the items (7, 11). When a respondent is confronted with an item which contains a familiar phrase to which he has given little thought, agreement may represent merely a passive and acquiescing response. A preponderance of responses indicating agreement to a large number of diverse items could thus result either from belief in all of the statements or from an acquiescent response set. Similarly, disagreement with an item may or may not reflect disbelief in its content; a person may be objecting merely to the style in which it is written.

The present study emphasizes the importance of acquiescence in relationships between the *F scale* and the Einstellung test. It is important to note, however, that Rokeach's (19) original findings were with reference to the *E scale* and Einstellung. Thus the conclusion that Rokeach's results are primarily attributable to acquiescence or response set may not be warranted by the present evidence. It would be valuable to repeat Rokeach's study substituting a different measure of anti-minority attitude, such as a social distance scale, to determine if this also correlates with problem solving behavior.

We do not mean to imply that the search for personality and cognitive correlates of attitudinal clusters is fruitless or that relationships between rigidity and attitudes are inconceivable. The work of Stern, Stein, and Bloom (20) indicates that when each end of an authoritarian scale is defined by both positively- and negatively-stated items, important relationships still appear, and in such a case it is possible to separate belief from response set.

Thus, if one end of an attitude continuum is defined by agreement with positively-stated items, as is the case with the F scale, it does not follow that the opposite end of the continuum can be designated by merely tallying disagreements to the same items. It should also be noted that even if the direction of item content had been controlled in constructing the F scale, the mathematical model implicit in a Likert scoring procedure requires a homogeneous scale. As Green (10) has pointed out, "the intercorrelations among the items are assumed to be due to a single common factor to which the items are all mutually related." Since there is evidence for multidimensionality in the F scale (e.g., 5), a revision more basic than just controlling for acquiescent response sets is probably necessary before it can justifiably be used for measuring an attitudinal dimension (10, 17).

Summary

This study is an attempt to appraise the role of a generalized tendency to agree with items independent of content in observed relationships between "authoritarianism" and Einstellung "rigidity." Subjects were given the Califórnia F scale, a reversed F scale, and a series of Einstellung water-jar problems. Subjects who used the long "rigid" solution in the Einstellung situation tended to agree with both F-scale and reversed F-scale items. These results suggested that both the Einstellung test and the F scale may be reflecting acquiescence and conformity. Also, the importance of defining both ends of a scale with positively-stated unidimensional items before drawing conclusions relating attitudes and personality was emphasized.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.
2. BASS, B. M. Authoritarianism or acquiescence? *J. abnorm. soc. Psychol.*, 1955, **51**, 611–623.
3. BROWN, R. W. A determinant of the relationship between rigidity and authoritarianism. *J. abnorm. soc. Psychol.*, 1953, **48**, 469–476.
4. COHN, T. S. The relation of the F scale to a response set to answer positively. *Amer. Psychologist*, 1953, **8**, 335. (Abstract)
5. CORDER, R. F. A factorial approach to anti-democratic attitudes. *Studies in higher education.* Lafayette, Ind.: Purdue Division of Educational Reference, 1954.
6. CRONBACH, L. J. Response sets and test validity. *Educ. psychol. Measmt*, 1946, **6**, 475–494.
7. CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt*, 1950, **10**, 3–31.
8. FRENCH, ELIZABETH G. Interrelation among some measures of rigidity under stress. *J. abnorm. soc. Psychol.*, 1955, **51**, 114–118.
9. GOODSTEIN, L. D. Intellectual rigidity and social attitudes. *J. abnorm. soc. Psychol.*, 1953, **48**, 345–353.
10. GREEN, B. F. Attitude measurement. In G. Lindzey (Ed.), *Handbook of social psychology.* Cambridge: Addison-Wesley, 1954. Vol. 1. Pp. 335–369.
11. GULLIKSEN, H. Intrinsic validity. *Amer. Psychologist*, 1950, **10**, 511–517.
12. JACKSON, D. N. Stability in resistance to field forces. Unpublished doctor's dissertation, Purdue Univer., 1955.
13. JACKSON, D. N., & MESSICK, S. J. A note on "ethnocentrism" and acquiescent response sets. *J. abnorm. soc. Psychol.*, 1957, **54**, 132–134.
14. LEAVITT, H. J., HAX, H., & ROCHE, J. H. "Authoritarianism" and agreement with things authoritative. *J. Psychol.*, 1955, **40**, 215–221.
15. LEVITT, E. E., & ZELEN, S. L. The validity of the Einstellung test as a measure of rigidity. *J. abnorm. soc. Psychol.*, 1953, **48**, 573–580.
16. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.
17. MESSICK, S. J. Some recent theoretical developments in multidimensional scaling. *Educ. psychol. Measmt*, 1956, **16**, 82–100.
18. PITCHER, BARBARA, & STACEY, C. L. Is Einstellung rigidity a general trait? *J. abnorm. soc. Psychol.*, 1954, **49**, 3–6.
19. ROKEACH, M. Generalized mental rigidity as a factor in ethnocentrism. *J. abnorm. soc. Psychol.*, 1948, **43**, 259–278.
20. STERN, G., STEIN, M., & BLOOM, B. *Methods in personality assessment.* Glencoe, Ill.: Free Press, 1956.

Received March 9, 1956

ANALITY, AGGRESSION, AND ACQUIESCENCE

WILLIAM RABINOWITZ

Bank Street College of Education[1]

FARBER (4) has recently presented evidence of a relationship between anality and political aggression as hypothesized in psychoanalytic theory. He found a significant correlation of .37 between two tests, the first a five-item measure of anality and the other a four-item measure of political aggression. An examination of the items in the two tests shows that the correlation may be a function of what Cronbach has termed "response set" (2, 3) rather than the result of a relationship between anality and aggression. On each of the anality items, the "Yes" option was keyed positively, and on each of the political aggression items, the "Agree" and "Strongly Agree" options were keyed positively. The keying of the items is such that scores on the two tests are influenced in the same way by any general tendency on the part of subjects to acquiesce, i.e., to say "Yes," "Agree," or "Strongly Agree," when responding to test items. There is ample evidence that response sets, such as acquiescence, are a reliable source of variance in test scores (2, 3), including tests of attitudes (1, 6), interests (5), and personality characteristics (5). In view of this evidence, there is reason to believe that Farber's results may stem from acquiescent tendencies on the part of subjects, rather than an inherent relationship between anality and aggression.

The most appropriate method of studying the anality-aggression relationship is to employ tests in which response sets cannot systematically influence scores. In the absence of such instruments, an alternative procedure is to add to each of Farber's tests an equal number of items which measure anality and political aggression, but which are keyed so that the "No" and "Disagree" options are positive. By examining the correlations among the original Farber tests and those newly constructed, it should be possible to determine whether the reported anality-aggression relation-

[1] Now at the Division of Teacher Education, Municipal Colleges of New York.

ship is an artifact of the way in which the test items were keyed.

Using the concept of anality as summarized by Farber, several new test items which seemed to embody this trait were constructed. Unlike Farber's items, however, the items were stated so that a "No" response would signify greater anality. The five items follow:

I usually lend my possessions to people who ask permission to use them. Yes No
If I'm doing something I enjoy, I often stay up late at night, sometimes so late that I don't get a full night's sleep. Yes No
I'd rather lead a life of excitement and adventure than one which is stable, predictable, and well-organized. Yes No
Some people think it's a good idea to keep a budget so that purchases can be carefully planned, but in my case I don't think I could stick to a budget for very long. Yes No
When I go to a restaurant I like to order foods which I have never tasted before. Yes No

Henceforth, Farber's anality test will be referred to as the A_1 scale, the author's anality test as the A_2 scale, and the combined 10 items as the A_T scale.

Following Farber's concept of political aggression, a test consisting of four items was built to measure this attribute. Here too the items were so stated that a "Disagree" response would signify greater aggression. The author's four items follow:

There is good reason to believe that the nations of the world can live in peace. Agree Disagree
The foreign policy of the United States should be based on the assumption that most other countries are our friends, not our enemies. Agree Disagree
Capital punishment should be abolished. Agree Disagree
It is unhealthy in a democracy to suppress minority viewpoints no matter how distasteful they may be. Agree Disagree

Henceforth, Farber's test of political aggression will be referred to as the P_1 scale, the author's test of political aggression as the P_2 scale, and the combined 8 items as the P_T scale.[2]

[2] In Farber's study each of the political aggression items had four response options: Agree Strongly, Agree Slightly, Disagree Slightly, and Disagree Strongly. The simpler procedure of using only an Agree and Disagree option for each item was used in this study, because this eliminated the influence of another response set, i.e., willingness to go to extremes in registering agreement or disagreement.

TABLE 1
CORRELATIONS OF A_1, A_2, P_1, AND P_2
(N = 143)

	A_1	A_2	P_1	P_2
A_1	+.30†	+.12	+.18*	−.04
A_2		+.32†	−.05	+.15*
P_1			+.53†	+.16*
P_2				+.25†

* $.01 < p < .05$.
† Reliability coefficients based on Kuder-Richardson Formula 20.

The A_T and P_T scales were administered to a group of 143 undergraduate college students (approximately 90 per cent female) who responded anonymously. Both scales were administered in a single session, the P_T scale immediately following the A_T scale. The responses given by the subjects were summarized in the following scores:

	Test	Way of Obtaining Score
Anality scores	A_1	Number of "Yes" responses
	A_2	Number of "No" responses
	A_T	Sum of above two scores
Political Aggression scores	P_1	Number of "Agree" responses
	P_2	Number of "Disagree" responses
	P_T	Sum of above two scores
Acquiescence scores	A_T	Number of "Yes" responses
	P_T	Number of "Agree" responses

On the assumption that anality and political aggression are more closely related than a response set to acquiesce, the correlation of A_T and P_T should be higher when scored for anality and political aggression than when scored for acquiescence. The actual correlations are as follows:

A_T vs. P_T (scored for anality and aggression) . . . $r = +.09$ (not significant)

A_T vs. P_T (scored for acquiescence) . . . $r = +.27$ (significant at the .01 level)

These correlations suggest that the reported anality-aggression relationship is a function of acquiescence. Further evidence to support this view appears in the intercorrelations of A_1, A_2, P_1, and P_2 as presented in Table 1. Both the correlation of A_1 with P_1, and A_2 with P_2 reach significance, though Farber reported a higher correlation between A_1 and P_1. Since these correlations involve tests whose items are scored in the same direction, it is impossible to tell whether they are a function of acquiescence or an anality-aggression relationship. This ambiguity is not present in the correlations of A_1 with P_2 and A_2 with P_1, since the items

of the tests involved in these correlations are scored in opposite directions. The obtained nonsignificant correlations suggest that the relationship between anality and political aggression is dependent on keying items in the same direction.

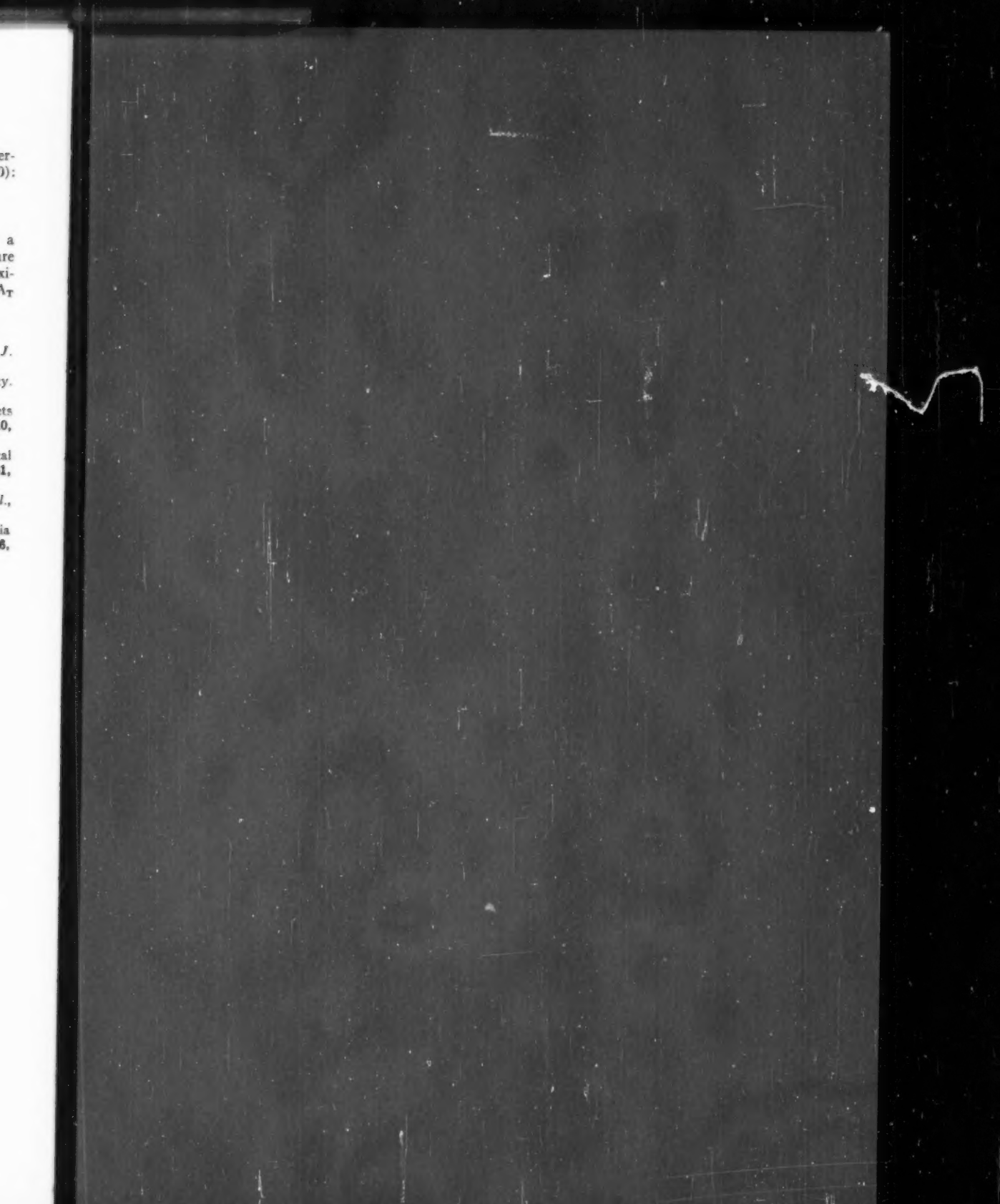
These results clearly suggest that a response set to acquiesce, not the reported anality-aggression relationship, adequately accounts for Farber's findings. A possible objection can be raised, however, to the scales developed by the author as inadequate measures of anality and political aggression. The low correlations between A_1 and A_2, and P_1 and P_2 indicate that Farber's scales and those developed in this study are not measuring the same attributes. The most obvious difference between the respective scales is that acquiescent tendencies are weighted differently in Farber's and the author's tests. The reliabilities reported in Table 1 indicate that A_2 and A_1 are about equally homogeneous. The author's test of anality is thus as internally consistent as is Farber's scale. However, since P_2 appears less reliable than P_1, Farber's test of political aggression seems the more adequate measure. Nevertheless, the total scales when scored for anality and political aggression are fairly homogeneous, as evidenced by their Kuder-Richardson reliability coefficients (Formula #20):

$$A_T \ldots r = .53$$

$$P_T \ldots r = .47$$

These reliabilities are high enough to permit a correlation of +.50 between A_T and P_T. The failure of the obtained correlation to approach this maximum cannot be attributed to unreliability of A_T and P_T due to the items added in this study.

REFERENCES

1. BASS, B. M. Authoritarianism or acquiescence? *J. abnorm. soc. Psychol.*, 1955, **51**, 616–623.
2. CRONBACH, L. J. Response sets and test validity. *Educ. psychol. Measmt*, 1946, **6**, 475–494.
3. CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt*, 1950, **10**, 3–31.
4. FARBER, M. L. The anal character and political aggression. *J. abnorm. soc. Psychol.*, 1955, **51**, 486–489.
5. LORGE, I. Gen-like: halo or reality. *Psychol. Bull.*, 1937, **34**, 545–546.
6. SHELLEY, H. P. Response set and the California attitude scales. *Educ. psychol. Measmt*, 1955, **16**, 63–67.

Received April 26, 1956